

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:30:40 ; Search time 9.67742 Seconds  
(without alignments)  
688.462 Million cell updates/sec

Title: US-09-863-901-22  
Perfect score: 258  
Sequence: 1 SGGSGGGGSGGSGGSGG.....SGSGGGGSGGSGGSLRS 50

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	258	100.0	50	ABB08752	GFP cloning interm
2	258	100.0	50	ABB08757	GFP peptide linker
3	258	100.0	477	ABB08634	GFP fusion protein
4	258	100.0	906	ABB08635	GFP fusion protein
5	212	82.2	468	ABB08633	GFP fusion protein
6	211	81.8	41	ABB08756	GFP peptide linker
7	170	65.9	738	AAW56163	New DNA sequence i
8	154	59.7	235	AAW05704	Glycine-rich repea
9	154	59.7	235	AAW79126	Epstein Barr Virus
10	154	59.7	261	AAW79137	FLGA Gly-ala inser

11	154	59.7	641	20	AAW28843	Epstein Barr Virus
12	154	59.7	641	21	AAW5856	Epstein Barr virus
13	154	59.7	641	22	AAW62332	EBV tethering prot
14	150	58.1	60	22	AAW77867	MHC class II H dom
15	150	58.1	60	22	AAW05183	Pain-relieving tar
16	148	57.4	265	19	AAW79128	FLGA insert stabil
17	147.5	57.2	532	23	AAW69423	Lung small cell ca
18	147.5	57.2	533	22	AAW05812	Human small cell l
19	147	57.0	56	23	ABP29010	Streptococcus poly
20	147	57.0	357	22	ABB29912	Peptide #2563 enco
21	147	57.0	357	22	ABB35090	Peptide #2596 enco
22	147	57.0	357	22	ABB20509	Protein #2508 enco
23	147	57.0	357	22	AAW55912	Human brain. expres
24	147	57.0	357	22	AAW68282	Human bone marrow
25	147	57.0	357	22	AAW16105	Peptide #2539 enco
26	147	57.0	357	22	AAW28596	Peptide #2633 enco
27	147	57.0	357	22	AAW03832	Peptide #2514 enco
28	147	57.0	357	23	ABG37823	Human peptide enco
29	146	56.6	399	22	ABG28641	Novel human diago
30	141.5	54.8	312	18	AAW18564	Novel fusion prote
31	140	54.3	40	21	AAW87573	Linker between CH3
32	140	54.3	610	22	ABW67896	Drosophila melanog
33	139.5	54.1	618	21	AAW56803	Human prostate can
34	139	53.9	358	21	AAW50939	Human adult skin c
35	139	53.9	369	21	AAW50940	Human adult skin c
36	139	53.9	387	21	AAW38324	Human secreted pro
37	139	53.9	479	21	AAW50941	Human adult skin c
38	137	53.1	283	22	ABB30839	Peptide #3490 enco
39	137	53.1	283	22	ABB36017	Peptide #3523 enco
40	137	53.1	283	22	ABB21422	Protein #3421 enco
41	137	53.1	283	22	AAW56810	Human brain expres
42	137	53.1	283	22	AAW69190	Human bone marrow
43	137	53.1	283	22	AAW17023	Peptide #3457 enco
44	137	53.1	283	22	AAW29514	Peptide #3551 enco
45	137	53.1	283	22	AAW04731	Peptide #3413 enco

## ALIGNMENTS

### RESULT 1

ABB08752  
ID ABB08752 standard; Protein; 50 AA.

XX ABB08752;

AC ABB08752;

XX 03-MAY-2002 (first entry)

DT DT

XX GFP cloning intermediate protein sequence 9.

DE GFP; aquorin; green fluorescent protein; photoprotein; bioluminescence;

KW Chemiluminescence Resonance Energy Transfer; CRET;

KW central nervous system; neural network.

XX Synthetic.

XX Key

PH Key

FT Region

FT Location/Qualifiers

FT 1..9

FT /note= "9 amino acid repeat region"

FT Misc-difference 1..45

FT /note= "region optionally deleted for 1-4 copies of the 9 amino acid repeats"

XX WO200192300-A2.

XX 06-DEC-2001.

XX 01-JUN-2001; 2001WO-EP07057.

XX 01-JUN-2000; 2000US-208314P.

XX 09-JUN-2000; 2000US-210526P.

XX 14-DEC-2000; 2000US-255111P.

PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CENT NAT RECH SCI.  
 XX  
 XX Baubet V, Le Mouelllic H, Brulet P;  
 XX WPI; 2002-139605/18;  
 XX N-PSDB; ABL41205.  
 XX  
 XX New fusion protein comprising a modified bioluminescent system with a  
 PT fluorescent molecule covalently linked with a photoprotein, useful for  
 PT monitoring calcium fluxes or for detecting electrical activity in a  
 PT group of neural cells  
 XX  
 XX Claim 24; Page 35; 61pp; English.  
 XX  
 XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy  
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence  
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule  
 CC covalently linked with a photoprotein. The fusion protein comprises the  
 CC formula: GFP - LINKER - AEQ  
 CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a  
 CC polypeptide of 4-63 amino acids.  
 CC The fusion protein is useful as a bioluminescent Ca<sup>2+</sup> reporter at the  
 CC single cell level. The fusion protein, composition or bioluminescent  
 CC system is useful for monitoring calcium fluxes in real time. This is  
 CC particularly useful for understanding the development, the plasticity and  
 CC the functioning of the central nervous system. The fusion protein  
 CC comprising the bioluminescent system is useful for detecting electrical  
 CC activity in a group of neural cells, for making it possible to complete  
 CC the phenotype study of mutants, or for observing the calcium activity in  
 CC a population of connected cells, for example in a neural network. The  
 CC present sequence is that of an intermediate sequence during cloning  
 CC manipulation of the GFP plasmid used in methods of the invention.  
 XX  
 XX Sequence 50 AA;  
 XX  
 XX Query Match 100.0%; Score 258; DB 23; Length 50;  
 XX Best Local Similarity 100.0%; Pred. No. 3.7e-17;  
 XX Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 50  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 50  
 RESULT 2  
 ABB08757  
 ID ABB08757 standard; peptide; 50 AA.  
 XX  
 XX ABB08757;  
 XX  
 XX 03-MAY-2002 (first entry)  
 XX  
 XX GFP peptide linker SEQ ID NO 22.  
 DE  
 DE GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;  
 KW Chemiluminescence Resonance Energy Transfer; CRET;  
 KW central nervous system; neural network.  
 XX  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FT FT Misc-difference 445 /note= "Encoded by AQT"  
 FT FT  
 XX  
 XX WO200192300-A2.  
 XX  
 XX 06-DEC-2001.  
 PD  
 PF 01-JUN-2001; 2001WO-EP07057.  
 XX  
 XX 01-JUN-2000; 2000US-208314P.  
 PR 09-JUN-2000; 2000US-210526P.  
 PR 14-DEC-2000; 2000US-255111P.  
 XX  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CENT NAT RECH SCI.

XX  
 PI Baubet V, Le Mouelllic H, Brulet P;  
 XX WPI; 2002-139605/18.  
 XX  
 XX New fusion protein comprising a modified bioluminescent system with a  
 PT fluorescent molecule covalently linked with a photoprotein, useful for  
 PT monitoring calcium fluxes or for detecting electrical activity in a  
 PT group of neural cells  
 XX  
 XX Claim 24; Page 35; 61pp; English.  
 XX  
 XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy  
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence  
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule  
 CC covalently linked with a photoprotein. The fusion protein comprises the  
 CC formula: GFP - LINKER - AEQ  
 CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a  
 CC polypeptide of 4-63 amino acids.  
 CC The fusion protein is useful as a bioluminescent Ca<sup>2+</sup> reporter at the  
 CC single cell level. The fusion protein, composition or bioluminescent  
 CC system is useful for monitoring calcium fluxes in real time. This is  
 CC particularly useful for understanding the development, the plasticity and  
 CC the functioning of the central nervous system. The fusion protein  
 CC comprising the bioluminescent system is useful for detecting electrical  
 CC activity in a group of neural cells, for making it possible to complete  
 CC the phenotype study of mutants, or for observing the calcium activity in  
 CC a population of connected cells, for example in a neural network. The  
 CC present sequence is that of a GFP peptide linker sequence of the  
 CC invention.  
 XX  
 XX Sequence 50 AA;  
 XX  
 XX Query Match 100.0%; Score 258; DB 23; Length 50;  
 XX Best Local Similarity 100.0%; Pred. No. 3.7e-17;  
 XX Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 50  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 50  
 RESULT 3  
 ABB08634  
 ID ABB08634 standard; Protein; 477 AA.  
 XX  
 XX ABB08634;  
 XX  
 XX 03-MAY-2002 (first entry)  
 XX  
 XX GFP fusion protein GSA SEQ ID NO 5.  
 DE  
 DE GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;  
 KW Chemiluminescence Resonance Energy Transfer; CRET;  
 KW central nervous system; neural network.  
 XX  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FT FT Misc-difference 445 /note= "Encoded by AQT"  
 FT FT  
 XX  
 XX WO200192300-A2.  
 XX  
 XX 06-DEC-2001.  
 PD  
 PF 01-JUN-2001; 2001WO-EP07057.  
 XX  
 XX 01-JUN-2000; 2000US-208314P.  
 PR 09-JUN-2000; 2000US-210526P.  
 PR 14-DEC-2000; 2000US-255111P.  
 XX  
 XX (INSP ) INST PASTEUR.





















## RESULT 9

```

US-09-010-928B-4
; Sequence 4, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Lewis, Cheryl Y
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;

```

STREET: 8110 GATEHOUSE RD. SUITE 500E  
CITY: FALLS CHURCH  
STATE: VIRGINIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 22042  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

```

; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-010-928B-4

```

Query Match	52.1%	Score	134.5;	DB	2;	Length	907;
Best Local Similarity	57.4%;	Pred. No.	5.1e-05;				
Matches	27;	Conservative	5;	Mismatches	14;	Indels	1;
Gaps							1;

RECEIVED 10

; Sequence 8, Application US/08770761A  
: Patent No. 5914502

/ PUBLICATION NO.: 2014303  
 / GENERAL INFORMATION:  
 / APPLICANT: Kovacevic, Steven  
 / APPLICANT: Otto, Keith A.  
 / APPLICANT: Rao, Ramachandra N.  
 / TITLE OF INVENTION: FUSION PROTEINS COMPRISING CELL CYCLE  
 / TITLE OF INVENTION: REGULATORY PROTEINS  
 / NUMBER OF SEQUENCES: 8  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Eli Lilly and Company  
 / STREET: Lilly Corporate Center/Patent Division



RESULT 13  
US-08-819-539-7  
; Sequence 7, Application US/08819539  
; Patent No. 5859324  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Zhong-Min  
; APPLICANT: Beer, Steven V.  
; TITLE OF INVENTION: Hypersensitive Response  
; TITLE OF INVENTION: Induced Resistance In Plants  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/819,539  
; FILING DATE: 17-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/475,775  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 14603/10050  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-819-539-7

Query Match 48.3%; Score 124.5; DB 2; Length 344;  
Best Local Similarity 40.6%; Pred. No. 0.00017;  
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps 2;  
QY 3 GSGSGGSGSGSGGQ-----SGSGSGGSGG-SGSGGSGSGGSG 43  
Db 141 GNGVGGANGAGGCGGLAEALQIEILQAGGGAGAGGGVGGAGDGGSGAG 200  
QY 44 GQSG 47  
Db 201 GAGG 204

RESULT 14  
US-09-030-270A-7  
; Sequence 7, Application US/09030270A  
; Patent No. 5977060  
; GENERAL INFORMATION:  
; APPLICANT: Zitter, Thomas A.  
; APPLICANT: Wei, Zhong-Min  
; TITLE OF INVENTION: INSECT CONTROL WITH A  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: P.O. Box 1051, Clinton Square

CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/030,270A  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,226  
FILING DATE: 28-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 19603/1521  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-030-270A-7

Query Match 48.3%; Score 124.5; DB 2; Length 344;  
Best Local Similarity 40.6%; Pred. No. 0.00017;  
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps 2;  
QY 3 GSGSGGSGSGSGGQ-----SGSGSGGSGG-SGSGGSGSGGSG 43  
Db 141 GNGVGGANGAGGCGGLAEALQIEILQAGGGAGAGGGVGGAGDGGSGAG 200  
QY 44 GQSG 47  
Db 201 GAGG 204

RESULT 15  
US-08-984-207-7  
; Sequence 7, Application US/08984207  
; Patent No. 6235974  
; GENERAL INFORMATION:  
; APPLICANT: Qiu, Dewen  
; APPLICANT: Wei, Zhong-Min  
; APPLICANT: Beer, Steven V.  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED  
; TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/984,207  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 60/033,230
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1201
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-984-207-7

Query Match      48.38; Score 124.5; DB 4; Length 344;
Best Local Similarity 40.68; Pred. No. 0.00017;
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps 2;

QY      3 GSGSGGSGSGSGSGGQ-----SGSGSGSGSGSGG--SGSGSGSGSGSGG 43
Db      141 GNGVGGANGAKGAGGGGGLAEALQETEQILQLGGGGAGAGAGGGVGGAGGADGGSGG 200

QY      44 GQSG 47
Db      201 GAGG 204

```

Search completed: June 18, 2003, 17:32:56  
Job time : 4.41536 secs











```

; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/296,801
; PRIOR FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly Gly tag
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be present
; OTHER INFORMATION: or absent
; US-10-161-165-3

```

Query Match	56.6%	Score 146;	DB 9;	Length 200;
Best Local Similarity	56.5%;	Pred. No. 3.3e-05;		
Matches 26;	Conservative	0;	Mismatches 20;	Indels 0;
			Gaps	0;

[illegible]

## RESULT 9

```

US-10-160-663-3
; Sequence 3, Application US/10160663
; Publication NO. US20030040001A1
; GENERAL INFORMATION:
; APPLICANT: Demo, Susan
; APPLICANT: Hitoshi, Yasumichi
; APPLICANT: Pearsall, Denise
; APPLICANT: Rigel Pharmaceutical, Incorporated
; TITLE OF INVENTION: LETM1: Modulators of Cellular Proliferation
; FILE REFERENCE: 021044-000920US
; CURRENT APPLICATION NUMBER: US/10/160,663
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/296,817
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/347,970
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly Gly tag
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be present
; OTHER INFORMATION: or absent
US-10-160-663-3

```

Query Match	56.6%	Score 146;	DB 9;	Length 200;
Best Local Similarity	56.5%;	Pred. No#	3.3e-05;	
Matches 26;	Conservative	0;	Mismatches 20;	Indels 0;
				Gaps 0;

**Qy**      2 GGSGGGQSGGSGSGSQSGGSGGGQSGSGSGGSQGSGGSGG 47  
          || || | | | | | | | | | | | | | | | | | | | |  
**Db**      1 GGG 46

## RESULT 10

US-10-071-838-15  
; Sequence 15, Application US/10071838

```

; Publication NO. US20030044814A1
; GENERAL INFORMATION:
; APPLICANT: Li, Jing
; APPLICANT: Powers, Scott
; APPLICANT: Xiang, Phil
; APPLICANT: Peng, Yue
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: PRC17: An Amplified Cancer Gene
; FILE REFERENCE: 018781-007610US
; CURRENT APPLICATION NUMBER: US/10/071,838
; CURRENT FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: US 60/267,615
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be pro
US-10-071-838-15

```

Query Match 56.6%; Score 146; DB 9; Length 200;  
Best Local Similarity 56.5%;  
Matches 26; Conservative 0; Mismatches 20; Indels  
Pred. No. 3.3e-05;

[illegible]

RESULT 11

```

US-10-094-417-25
/ Sequence 25, Application US/10094417
/ Publication No. US20030045685A1
/ GENERAL INFORMATION:
/ APPLICANT: Tian, Hui
/ APPLICANT: Zhao, Jiagang
/ APPLICANT: Chen, Jin-Long
/ APPLICANT: Cutler, Gene
/ APPLICANT: Tularik Inc.
/ TITLE OF INVENTION: No. US20030045685A1el Receptors
/ FILE REFERENCE: 018781-0081100S
/ CURRENT APPLICATION NUMBER: US/10/094, 417
/ CURRENT FILING DATE: 2002-06-10
/ PRIOR APPLICATION NUMBER: US 09/802, 803
/ PRIOR FILING DATE: 2001-03-09
/ PRIOR APPLICATION NUMBER: US 60/276, 649
/ PRIOR FILING DATE: 2001-03-16
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 25
/ LENGTH: 200
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence
/ OTHER INFORMATION: flexible linker
/ FEATURE:
/ NAME/KEY: MOD_RES
/ LOCATION: (6)..(200)
/ OTHER INFORMATION: Gly at positions 6-200 may be pr
US-10-094-417-25

```

Query Match 56.6%; Score 146; DB 9; Length 200;  
Best Local Similarity 56.5%; Pred. No. 3.3e-05;  
Matches 26; Conservative 0; Mismatches 20; Indels



Query Match	56.6%	Score 145;	DB 9;	Length 200;
Best Local Similarity	56.5%	Pred. No. 3.e-05;		
Matches	26;	Conservative 0;	Mismatches 20;	Indels 0;
			Gaps	0;

Search completed: June 18, 2003, 17:44:07  
Job time : 7.07211 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 17:30:43 ; Search time 41.9355 Seconds  
(without alignments)  
768.720 Million cell updates/sec

Title: US-09-863-901-22

Perfect score: 258  
Sequence: 1 SGGSGGGGSGSGGSGG.....SGSGGGGSGSGGSGGSLRS 50

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main.\*  
1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US103\_COMB.pep.\*  
28: /cgn2\_6/ptodata/1/paa/US104\_COMB.pep.\*  
29: /cgn2\_6/ptodata/1/paa/US105\_COMB.pep.\*  
30: /cgn2\_6/ptodata/1/paa/US106\_COMB.pep.\*  
31: /cgn2\_6/ptodata/1/paa/US107\_COMB.pep.\*  
32: /cgn2\_6/ptodata/1/paa/US108\_COMB.pep.\*  
33: /cgn2\_6/ptodata/1/paa/US109\_COMB.pep.\*  
34: /cgn2\_6/ptodata/1/paa/US110\_COMB.pep.\*  
35: /cgn2\_6/ptodata/1/paa/US111\_COMB.pep.\*  
36: /cgn2\_6/ptodata/1/paa/US112\_COMB.pep.\*  
37: /cgn2\_6/ptodata/1/paa/US113\_COMB.pep.\*  
38: /cgn2\_6/ptodata/1/paa/US114\_COMB.pep.\*  
39: /cgn2\_6/ptodata/1/paa/US115\_COMB.pep.\*  
40: /cgn2\_6/ptodata/1/paa/US116\_COMB.pep.\*  
41: /cgn2\_6/ptodata/1/paa/US117\_COMB.pep.\*  
42: /cgn2\_6/ptodata/1/paa/US118\_COMB.pep.\*  
43: /cgn2\_6/ptodata/1/paa/US119\_COMB.pep.\*  
44: /cgn2\_6/ptodata/1/paa/US120\_COMB.pep.\*  
45: /cgn2\_6/ptodata/1/paa/US121\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258	100.0	50	22	US-09-863-901-22
2	258	100.0	477	22	US-09-863-901-5
3	258	100.0	906	22	US-09-863-901-6
4	235	91.1	45	22	US-09-863-901-25
5	212	82.2	468	22	US-09-863-901-4
6	211	81.8	41	22	US-09-863-901-21

7	170	65.9	300	21	US-09-708-427-19850	Sequence 19850, A
8	170	65.9	302	21	US-09-708-427-19849	Sequence 19849, A
9	168	65.1	592	27	US-60-360-039-17534	Sequence 17534, A
10	163	63.2	1307	21	US-09-708-427-4493	Sequence 4493, Ap
11	163	63.2	1349	21	US-09-708-427-4492	Sequence 4492, Ap
12	163	63.2	1422	21	US-09-708-427-4491	Sequence 4491, Ap
13	155.5	60.3	138	26	US-10-219-999-45027	Sequence 45027, A
14	154	59.7	235	9	US-08-529-190A-1	Sequence 1, Appl
15	154	59.7	235	13	US-08-970-900-1	Sequence 1, Appl
16	154	59.7	266	13	US-08-970-900-5	Sequence 5, Appl
17	154	59.7	267	11	US-08-733-369A-106	Sequence 106, App
18	154	59.7	603	3	US-07-865-030-1	Sequence 1, Appl
19	154	59.7	641	18	US-09-410-399-4	Sequence 4, Appl
20	154	59.7	641	21	US-09-791-537-91237	Sequence 91237, A
21	154	59.7	641	25	US-10-138-098-52	Sequence 52, Appl
22	153	59.3	1115	21	US-09-708-427-4736	Sequence 4736, Ap
23	153	59.3	1145	21	US-09-708-427-4735	Sequence 4735, Ap
24	153	59.3	1218	21	US-09-708-427-4734	Sequence 4734, Ap
25	150	58.1	50	18	US-09-421-971-73	Sequence 73, Appl
26	150	58.1	50	23	US-09-949-039-32	Sequence 32, Appl
27	150	58.1	54	18	US-09-421-971-64	Sequence 64, Appl
28	150	58.1	55	18	US-09-421-971-74	Sequence 74, Appl
29	150	58.1	59	1	PCT-US00-41224-13	Sequence 13, Appl
30	150	58.1	59	18	US-09-421-971-65	Sequence 65, Appl
31	150	58.1	60	18	US-09-421-971-75	Sequence 75, Appl
32	150	58.1	60	22	US-09-833-203-18	Sequence 18, Appl
33	150	58.1	100	1	PCT-US00-27794B-4	Sequence 4, Appl
34	150	58.1	100	18	US-09-411-067C-4	Sequence 4, Appl
35	148.5	57.6	171	21	US-09-708-427-6562	Sequence 6562, Ap
36	147.5	57.2	532	22	US-09-833-790-428	Sequence 428, App
37	147.5	57.2	533	18	US-09-489-101A-22	Sequence 22, Appl
38	147.5	57.2	533	21	US-09-791-537-142569	Sequence 142569, A
39	147	57.0	175	18	US-09-417-507-43604	Sequence 43604, A
40	147	57.0	188	18	US-09-417-507-22219	Sequence 22219, A
41	147	57.0	201	1	PCT-US01-00358-5	Sequence 5, Appl
42	147	57.0	201	1	PCT-US99-17885-11	Sequence 11, Appl
43	147	57.0	201	14	US-09-054-231-35	Sequence 35, Appl
44	147	57.0	201	15	US-09-130-858-11	Sequence 11, Appl
45	147	57.0	201	18	US-09-470-850-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1  
US-09-863-901-22  
; Sequence 22, Application US/09863901  
; GENERAL INFORMATION:  
; APPLICANT: BAUBET, VALERIE  
; APPLICANT: LE MOUILLIC, HERVE  
; APPLICANT: BRULET, PHILIPPE  
; TITLE OF INVENTION: CHIMERIC GFP-AEQUORIN AS BIOLUMINESCENT Ca++ REPORTERS  
; FILE REFERENCE: 03495-0207-00000  
; CURRENT APPLICATION NUMBER: US/09/863,901  
; CURRENT FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: 60/208,314  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: 60/210,526  
; PRIOR FILING DATE: 2000-06-06  
; PRIOR APPLICATION NUMBER: 60/255,111  
; PRIOR FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 50  
; TYPE: PPT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Peptide  
; OTHER INFORMATION: sequence of linker  
US-09-863-901-22

ORGANISM: *Aequorea victoria*  
S-09-863-901-4

Query Match	82.2%	Score 212;	DB 22;	Length 468;
-------------	-------	------------	--------	-------------





```

; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4491
; LENGTH: 1422
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1422
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..1422
; OTHER INFORMATION: Ceres Seq. ID 1812441
US-09-708-427-4491

Query Match 63.2%; Score 163; DB 21; Length 1422;
Best Local Similarity 61.7%; Pred. No. 6.6e-06;
Matches 29; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

Qy 1 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSG 47
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 SGGDGGGPGSGDGGGPGSGDGGGPGSGDGGGPGSGDGGGPG 472

RESULT 13
US-10-219-999-45027
; Sequence 45027, Application US/10219999
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)C
; CURRENT APPLICATION NUMBER: US/10/219,999
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/324,109
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/312,544
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 63520
; SEQ ID NO 45027
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Zea mays
US-10-219-999-45027

Query Match 60.3%; Score 155.5; DB 26; Length 138;
Best Local Similarity 59.8%; Pred. No. 3.8e-06;
Matches 30; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

Qy 2 GGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGG 44
    ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 GGGSGGGGSGGGSGGGG-SGGGGSGGGSGGGSGGGSGGGSG 49

RESULT 14
US-08-529-190A-1
; Sequence 1, Application US/08529190A
; GENERAL INFORMATION:
; APPLICANT: Masucci, Maria G.
; TITLE OF INVENTION: GLYCINE-CONTAINING SEQUENCES
; TITLE OF INVENTION: CONFERRING INVISIBILITY TO THE IMMUNE SYSTEM
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti
; STREET: 75 State Street
; CITY: Boston

```







**Db**

415 GEGGPGSGDGECCGPSSGGDEGGPSCADGEGGPSGGDGEGGPSG 457







**C;Keywords:** coiled coil; heterotetramer; intermediate filament; polymorphism





DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sericin-1 (Silk gum protein 1) (Fragment).  
 GN SER1 OR SER-1.  
 OS Galleria mellonella (Wax moth).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
 OC Ditrysia; Pyraloidea; Pyralidae; Galleriinae; Galleriinae; Galleria.  
 ON NCBI\_TaxID=7137;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Middle silk gland;  
 RA Zurovec M., Sehna F., Scheller K., Kumaran A.K.;  
 RT "Silk gland specific cDNAs from Galleria mellonella L.";  
 RL Insect Biochem. Mol. Biol. 22:55-67(1992).  
 CC -1- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING.  
 CC ACTS AS A CEMENT BY STICKING SILK THREADS TOGETHER.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG)  
 CC SECTION OF SILK GLANDS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AF095241; AAC79078.1;  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen; 1.  
 DR SIKK.  
 FT NON\_TER 1 1  
 FT SEQUENCE 115 AA; 9161 MW; FC678FCCA95E4999 CRC64;  
 SQ  
 Query Match 62.4%; Score 161; DB 1; Length 115;  
 Best Local Similarity 66.0%; Pred. No. 3.6e-06;  
 Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
 OY 1 SGGSS 50  
 DB 36 SGGSS 85  
 -----  
 RESULT 3  
 ID EBNI\_EBV  
 AC P03211; STANDARD; PRT; 641 AA.  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE EBNA-1 nuclear protein.  
 GN BKRF1.  
 OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 ON NCBI\_TaxID=10377;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=84270667; PubMed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,  
 RA Tufnell P.S., Warrall B.G.;  
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";  
 RL Nature 310:207-211(1984).  
 RN [2]  
 RP SEQUENCE OF 1-26 FROM N.A.  
 RX MEDLINE=86259739; PubMed=3460083;  
 RA Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;  
 RA "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear  
 RT proteins: a probable transcriptional initiation site";  
 RT Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).

RN [3]  
 RP SUBCELLULAR LOCATION.  
 RA MEDLINE=90266473; PubMed=21611150;  
 RA Petti L., Sample C., Kieff E.;  
 RT "Subnuclear localization and phosphorylation of Epstein-Barr virus  
 RT latent infection nuclear proteins";  
 RL Virology 176:563-574(1990).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.  
 RA MEDLINE=96006523; PubMed=7553871;  
 RA Hochkarev A., Barwell J.A., Pfuetzner R.A., Furey W.F. Jr.,  
 RA Edwards A.M., Frappier L.;  
 RT "Crystal structure of the DNA-binding domain of the Epstein-Barr  
 RT virus origin-binding protein EBNA 1";  
 RL Cell 83:39-46(1995).  
 CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE  
 CC MAINTENANCE REPLICATION OF EBV EPISOME. TRANSCRIPTIVATING FACTOR  
 CC FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM, SOMEWHAT  
 CC ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED  
 CC WITH THE NUCLEAR MATRIX.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; V01555; CAA24816.1;  
 DR EMBL; M13941; AAA45889.1;  
 DR PIR; A03773; Q0BE31.  
 DR PIR; S33021; S33021.  
 DR PDB; 1VHL; 23-DEC-96.  
 DR TRANSFAC; T00211;  
 DR InterPro; IPR004186; EBNA1.  
 DR Pfam; PF02905; EBNA1; 1.  
 DR KW Nuclear protein; DNA-binding; Transcription regulation; Activator;  
 FT 3D-structure. 87 352 GLY/ALA-RICH.  
 FT DOMAIN 87 352  
 FT SEQUENCE 641 AA; 58427 MW; 4D161653E16FC341 CRC64;  
 SQ  
 Query Match 59.7%; Score 154; DB 1; Length 641;  
 Best Local Similarity 54.3%; Pred. No. 3.9e-05;  
 Matches 25; Conservative 13; Mismatches 8; Indels 0; Gaps 0;  
 OY 2 GSG 47  
 DB 146 GGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 191  
 -----  
 RESULT 4  
 ID ZIC2\_HUMAN  
 AC Q95409; Q9H309; STANDARD; PRT; 532 AA.  
 DT 30-JUN-2000 (Rel. 39, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Zinc finger protein ZIC2 (Zinc finger protein of the cerebellum 2).  
 GN ZIC2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANT HPES POLY-ALA INSERTION.  
 RX MEDLINE=98442655; PubMed=9771712;  
 RA Brown S.A., Warburton D., Brown L.Y., Yu C.Y., Roeder E.R.,  
 RA Stengel-Rutkowski S., Hennekam R.C., Muenke M.;  
 RT "Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila  
 RT odd-paired";  
 RL Nat. Genet. 20:180-183(1998).













```

RESULT 12
K2C5_BOVIN STANDARD; PRT; 166 AA.
ID AC P04262;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II cytoskeletal 68 kDa, component IB (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
RN [1]
RX SEQUENCE FROM N.A.
MEDLINE=85128114; PubMed=6084625;
RA Jorcano J.L., Franz J.K., Franke W.W.;
RT "Amino acid sequence diversity between bovine epidermal cytokeratin
RT polypeptides of the basic (type II) subfamily as determined from cDNA
RT clones."
RL Differentiation 28:155-163(1984).
CC -I- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC -I- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; K03534; AAA30601.1; -.
DR FIR; A02948; KRBO2B.
DR InterPro; IPRO01664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
DR Intermediate filament; Coiled coil; Keratin.
KW NON_TER 1 1
FT DOMAIN <1 41 ROD.
FT DOMAIN 42 166 TAIL.
FT DOMAIN <1 1 COIL_2B.
SQ SEQUENCE 166 AA; 15505 MW; 4BFE495A7C6B4BE0 CRC64;
      Query Match          55.2%; Score 142.5; DB 1; Length 166;
      Best Local Similarity 66.7%; Pred. No. 9.6e-05;
      Matches 36; Conservative 2; Mismatches 9; Indels 7; Gaps 3;

QY 2 GGSGGGGSGGGSGGGSGGS--GGQSGSGSGS-----GGQSG-GSGSGGQSGL 48
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 87 GGSGFGGGSGIGSGSGFGGGSGFGGGSGFGGGSGFGGGSGFGGGSGGRSV 140

RESULT 13
LORL_MOUSE STANDARD; PRT; 481 AA.
ID LORL_MOUSE
AC F18165;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lorcrin.
OS Mus musculus (Mouse).
OC OS
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RX SEQUENCE FROM N.A.
MEDLINE=90275605; PubMed=2190691;
RA Mehrl T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.
```





GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:36:34 ; Search time 9.7723 Seconds  
(without alignments)  
1054.241 Million cell updates/sec

Title: US-09-863-901-22

Perfect score: 258

Sequence: 1 SGSGSGGSGSGSGGSG.....SGSGGSGSGSGGSGGSLRS 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	170	65.9	302	10 Q9SL09	Q9SL09 arabidopsis
2	170	65.9	738	5 Q02402	Q02402 pinctada fu
3	169	65.5	125	5 Q964C1	Q964C1 encephalito
4	168	65.1	592	16 Q9PF60	Q9PF60 xylella fas
5	166	64.3	102	5 Q964C0	Q964C0 encephalito
6	163	63.2	1422	10 Q9ZUR3	Q9ZUR3 arabidopsis
7	163	63.2	1745	10 Q9MAK1	Q9MAK1 arabidopsis
8	161.5	62.6	100	5 Q964C3	Q964C3 encephalito
9	161.5	62.6	117	5 Q964C2	Q964C2 encephalito
10	161.5	62.6	121	5 Q964C5	Q964C5 encephalito
11	161.5	62.6	138	5 Q964C4	Q964C4 encephalito
12	159.5	61.8	207	10 Q43522	Q43522 lycopersico
13	154	59.7	786	6 O18740	O18740 canis fami
14	150	59.3	1218	10 Q9XIH2	Q9XIH2 arabidopsis
15	150	58.1	251	3 Q9UVE7	Q9UVE7 yarrowia li
16	149	57.8	157	5 Q9GQX8	Q9GQX8 plasmodium

17	148.5	57.6	171	10	004339	004339 arabidopsis
18	147.5	57.2	87	10	Q9LRIS	Q9LRIS pisum sativ
19	146	56.6	291	10	Q39337	Q39337 brassica na
20	146	56.6	370	5	Q9BPK8	Q9BPK8 plasmodium
21	146	56.6	370	5	Q9BPK6	Q9BPK6 plasmodium
22	145.5	56.4	118	10	Q949R3	Q949R3 arabidopsis
23	145	56.2	277	5	Q9GQF0	Q9GQF0 plasmodium
24	145	56.2	302	5	Q9GQL9	Q9GQL9 plasmodium
25	145	56.2	302	5	Q9GQL8	Q9GQL8 plasmodium
26	145	56.2	317	5	Q9BH54	Q9BH54 plasmodium
27	144.5	56.0	221	10	065514	065514 arabidopsis
28	144	55.8	561	11	Q9CXH6	Q9CXH6 mus musculu
29	143.5	55.6	208	5	Q25949	Q25949 plasmodium
30	143.5	55.6	718	12	Q91TR1	Q91TR1 tupiaia herp
31	143	55.4	588	12	Q9QEK6	Q9QEK6 cynomolgus
32	142	55.0	486	11	Q8R0T9	Q8R0T9 mus musculu
33	142	55.0	698	12	Q8QXK8	Q8QXK8 ectocarpus
34	141.5	54.8	619	12	Q9IFQ9	Q9IFQ9 cynomolgus
35	141	54.7	139	5	Q9GQX9	Q9GQX9 plasmodium
36	141	54.7	208	5	Q9U0A0	Q9U0A0 plasmodium
37	141	54.7	396	10	065450	065450 arabidopsis
38	140	54.3	610	5	Q9V5V8	Q9V5V8 drosophila
39	139.5	54.1	637	11	Q9D2K8	Q9D2K8 mus musculu
40	139	53.9	185	10	Q948R3	Q948R3 oryza sativ
41	138.5	53.7	271	10	Q08529	Q08529 nicotiana t
42	138.5	53.7	588	12	Q9IPQ8	Q9IPQ8 cynomolgus
43	138	53.5	164	5	Q9BJO5	Q9BJO5 plasmodium
44	138	53.5	484	16	053394	053394 mycobacteri
45	137.5	53.3	175	10	Q9LSN6	Q9LSN6 arabidopsis

## ALIGNMENTS

RESULT 1

ID	Q9SL09	PRELIMINARY;	PRT;	302 AA.
AC	Q9SL09;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	AT2G05580 protein.			
GN	AT2G05580.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RX	MEDLINE=20083487; PubMed=10617197;			
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,			
RA	Fujii C.Y., Mason T.M., Bowman C.B., Barnstead M.E., Feldblyum T.V.,			
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,			
RA	Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,			
RA	Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,			
RA	Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,			
RA	Salzberg S.L., Fraser C.M., Venter J.C.;			
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis			
RL	thaliana";			
RL	Nature 402:761-768(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RA	Lin X.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC006220; AAD24649.1; -.			
DR	InterPro; IPR002952; Eggshell.			
DR	PRINTS; IPR000817; PRION.			
DR	PRINTS; PR01228; EGGSHLL.			
DR	PRINTS; PR00341; PRION.			
SQ	SEQUENCE 302 AA; 26008 MW; E72A73C55825891E CRC64;			



RA Xiao L., Li L., Visvesvara G.S., Moura H., Didier E.

RA Fraser C.M., Venter J.C.;





```

OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. UC82B; TISSUE=FRUIT;
RX MEDLINE=97201476; PubMed=9049262;
RA Santino C.G., Stanford G.L., Conner T.W.;
RT "Developmental and transgenic analysis of two tomato fruit enhanced
RL genes."; Biol. 33:405-416(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. UC82B; TISSUE=FRUIT;
RA Connor T.W.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X95262; CAAG4559.1; -.
DR HSSP; P24337; IHYP.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; tryr_alpha_aml; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 207 AA; 18265 MW; A50FE71F8ED4C4D9 CRC64;

Query Match 61.8%; Score 159.5; DB 10; Length 207;
Best Local Similarity 67.3%; Pred. No. 1.5e-07;
Matches 33; Conservative 1; Mismatches 12; Indels 3; Gaps 1;

Oy 2 GSGSGGSGGSGSGG---GSGSGGSGGSGGSGGSGGSGGSGGSGGSGG 47
Dy 54 GSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 102

RESULT 13
O18740
ID O18740 PRELIMINARY; PRT; 786 AA.
AC O18740;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Keratin.
GN KRT9.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Lacharme P., Hitt C., Jouquand S., Priat C., Galibert F.;
RT "Identification and analysis of the dog keratin 9 (KRT9) gene.";
RL Anim. Genet. 9:173-178(1998).
DR EMBL; AF000949; AAC26971.1; -.
DR HSSP; P10968; 2CWG.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; UNKNOWN_1.
SQ SEQUENCE 786 AA; 76355 MW; 272AB5425DD09535 CRC64;

Query Match 59.7%; Score 154; DB 6; Length 786;
Best Local Similarity 49.3%; Pred. No. 1.6e-06;
Matches 36; Conservative 3; Mismatches 8; Indels 26; Gaps 2;

Oy 1 SGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 37
Dy 661 SGGSSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 720

Oy 38 ---GSGSGSGGSGG 47
Dy 721 SYGGSGSGGSGGSGG 733

RESULT 14
O9XIH2

```

```

ID O9XIH2 PRELIMINARY; PRT; 1218 AA.
AC O9XIH2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE At2g16180 protein.
GN AT2G16180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Omayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RL thaliana";
RN Nature 402:761-768(1999).
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007134; AAD26959.1; -.
DR InterPro; IPR002106; AATRNA_ligaseII.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
SQ SEQUENCE 1218 AA; 136382 MW; 4736D7FCCB72245A CRC64;

Query Match 59.3%; Score 153; DB 10; Length 1218;
Best Local Similarity 65.1%; Pred. No. 3e-06;
Matches 28; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 5 GSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSG 47
Dy 415 GEGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSG 457

RESULT 15
O9UVE7
ID O9UVE7 PRELIMINARY; PRT; 251 AA.
AC O9UVE7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 21.8 kDa protein (fragment).
GN 45C.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 20460; TISSUE=MYCELIUM;
RA Sanchez M., Dominguez A.;
RT "Gene order in a 10275 bp fragment from Yarrowia lipolytica including
RL YLURA5 and YLSEC65 adjacent genes conserved in four yeast species.";
DR EMBL; AJ006754; CAB55339.1; -.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHELL.
DR PRINTS; PR01574; TUBBYPROTEIN.
DR Hypothetical protein.

```

Search completed: June 17, 2003, 16:43:27  
Job time : 12.7723 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:30:40 ; Search time 92.3226 Seconds  
(without alignments)  
688.462 Million cell updates/sec

Title: US-09-863-901-5

Perfect score: 2560

Sequence: 1 MSKGBELFGVPLVLDG.....LGFVYMDPACELKYGAVP 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2560	100.0	477	23	AB08634
2	2560	100.0	906	23	GFP fusion protein
3	2498.5	97.6	468	23	AB08633
4	2395.5	93.6	450	23	AB08632
5	2344	91.6	441	23	AB08631
6	2287.5	89.4	432	23	AB08630
7	1353	52.9	655	23	AB076672
8	1295	50.6	719	19	AAW85012
9	1295	50.6	719	21	AAW70779
10	1293.5	50.5	356	23	AB08620

11	1292.5	50.5	631	19	AAW85006
12	1292.5	50.5	631	21	AAW70778
13	1291.5	50.4	1386	21	AAW85573
14	1290.5	50.4	890	21	AAW22938
15	1288.5	50.3	294	21	AAW22860
16	1288.5	50.3	294	21	AAW79638
17	1286	50.2	1070	18	AAW17789
18	1283.5	50.1	633	19	AAW85009
19	1283	50.1	459	21	AAW22936
20	1278.5	49.9	635	21	AAW85034
21	1278.5	49.9	635	21	AAW70781
22	1277	49.9	501	18	AAW31879
23	1276.5	49.9	1171	19	AAW85037
24	1274	49.8	250	22	AAW86492
25	1272	49.7	558	19	AAW48662
26	1271.5	49.7	1090	21	AAW85576
27	1271	49.6	397	20	AAW42176
28	1271	49.6	429	20	AAW42175
29	1271	49.6	501	18	AAW31877
30	1271	49.6	501	18	AAW31878
31	1271	49.6	514	18	AAW31876
32	1270	49.6	238	18	AAW22101
33	1270	49.6	238	20	AAW96328
34	1270	49.6	247	20	AAW96329
35	1269.5	49.6	359	20	AAW42179
36	1269.5	49.6	391	20	AAW42180
37	1269.5	49.6	391	20	AAW42178
38	1269.5	49.6	997	19	AAW85032
39	1269	49.6	607	19	AAW85010
40	1269	49.6	783	21	AAW22939
41	1268	49.5	248	23	AAW86319
42	1268	49.5	403	20	AAW42177
43	1268	49.5	842	19	AAW85008
44	1267	49.5	605	19	AAW85007
45	1267	49.5	727	19	AAW85041

## ALIGNMENTS

### RESULT 1

AB08634

ID AB08634 standard; Protein; 477 AA.

XX AC AB08634;

XX DT 03-MAY-2002 (first entry)

XX DE GFP fusion protein G5A SEQ ID NO 5.

XX KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;

XX KW Chemiluminescence Resonance Energy Transfer; CRET;

XX KW central nervous system; neural network.

XX OS Synthetic.

XX Key Location/Qualifiers

XX FT Misc-difference 445

XX FT /note= "Encoded by AQT"

XX FN WO200192300-A2.

XX PD 06-DEC-2001.

XX PF 01-JUN-2001; 2001WO-EP07057.

XX PF 01-JUN-2000; 2000US-208314P.

XX PR 09-JUN-2000; 2000US-210526P.

XX PR 14-DEC-2000; 2000US-255111P.

XX PA (INSP ) INST PASTEUR.

XX PA (CNRS ) CENT NAT RECH SCI.

XX

Erk1-green floures  
EGFP-Erk1 fusion p  
Hs-UNC-53/3 fragm  
GFP-HSC70 fusion p  
GFP-DEVD-annexin I  
Caspase-3 biosenso  
Green fluorescent  
Jnk1-green floures  
GFP-HSP27 fusion p  
EGFP-VASP fusion p  
GFP variants S65T  
Green fluorescent  
GR fusion protein  
Hs-UNC-53/1 fragm  
EGFP/DRM fusion pr  
EGFP/DRM fusion pr  
GFP variants S65C  
GFP variants S65C  
Aequorea victoria  
Humanised green fl  
EGFP/DRM fusion pr  
EGFP/DRM fusion pr  
Green fluorescent  
p38-green flouresc  
GFP-HSP1 fusion pr  
Jellyfish green fl  
EGFP/DRM fusion pr  
Grk5-green floures  
Erk2-green floures  
Green fluorescent



Db 490 VTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTLV 549  
 Qy 121 NRIELKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGVSQVLAD 180  
 Db 550 NRIELKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGVSQVLAD 609  
 Qy 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240  
 Db 610 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 669  
 Qy 241 GSGSGGSG 300  
 Db 670 GSGSGGSG 729  
 Qy 301 IGRHKHMFNFDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHKDAVEAFPGGAGMK 360  
 Db 730 IGRHKHMFNFDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHKDAVEAFPGGAGMK 789  
 Qy 361 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWDALFDIVDKONGAITLDENKAY 420  
 Db 790 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWDALFDIVDKONGAITLDENKAY 849  
 Qy 421 TKAAGIIQSSDCEETFRVCDIDESGQLDVDEMTROHLGFWYTMDPACEKLYGGAVP 477  
 Db 850 TKAAGIIQSSDCEETFRVCDIDESGQLDVDEMTROHLGFWYTMDPACEKLYGGAVP 906

RESULT 3  
 ABB08633  
 ID ABB08633 standard; Protein: 468 AA.  
 AC ABB08633;  
 XX  
 DT 03-MAY-2002 (first entry)  
 XX  
 DE GFP fusion protein G4A SEQ ID NO 4.  
 XX  
 KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;  
 KW Chemiluminescence Resonance Energy Transfer; CRET;  
 KW central nervous system; neural network.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 401 /note= "Encoded by QGA"  
 XX  
 PN WO200192300-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 PF 01-JUN-2001; 2001WO-EP07057.  
 XX  
 PR 01-JUN-2000; 2000US-208314P.  
 PR 09-JUN-2000; 2000US-210526P.  
 PR 14-DEC-2000; 2000US-255111P.  
 XX  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CENT NAT RECH SCI.  
 XX  
 XX Baubet V, Le Mouellie H, Brulet P;  
 XX WPI; 2002-139605/18.  
 DR N-PSDB; ABA97888.  
 XX  
 PT New fusion protein comprising a modified bioluminescent system with a  
 PT fluorescent molecule covalently linked with a photoprotein; useful for  
 PT monitoring calcium fluxes or for detecting electrical activity in a  
 PT group of neural cells  
 XX  
 PS Claim 5; Page 31; 61pp; English.  
 XX  
 XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy

CC transfer from aequorin to green fluorescent protein by Chemiluminescence  
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule  
 CC covalently linked with a photoprotein. The fusion protein comprises the  
 CC formula: GFP - LINKER - AEO  
 CC GFP - green fluorescent protein; AEO - aequorin; and LINKER - a  
 CC polypeptide of 4-63 amino acids.  
 CC The fusion protein is useful as a bioluminescent Ca<sup>2+</sup> reporter at the  
 CC single cell level. The fusion protein, composition or bioluminescent  
 CC system is useful for monitoring calcium fluxes in real time. This is  
 CC particularly useful for understanding the development, the plasticity and  
 CC the functioning of the central nervous system. The fusion protein  
 CC comprising the bioluminescent system is useful for detecting electrical  
 CC activity in a group of neural cells, for making it possible to complete  
 CC the phenotype study of mutants, or for observing the calcium activity in  
 CC a population of connected cells, for example in a neural network.  
 XX  
 SQ Sequence 468 AA;  
 Query Match 97.6%; Score 2498.5; DB 23; Length 468;  
 Best Local Similarity 98.1%; Pred. No. 2.1e-202;  
 Matches 468; Conservative 0; Mismatches 0; Indels 9; Gaps 1;  
 Qy 1 MSKEELFTGVVPIVLVDGVDNGHKFSVSGEGEDATYGLTKLFTCTGKLPVWPPTL 60  
 Db 1 MSKEELFTGVVPIVLVDGVDNGHKFSVSGEGEDATYGLTKLFTCTGKLPVWPPTL 60  
 Qy 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTLV 120  
 Db 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTLV 120  
 Qy 121 NRIELKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGVSQVLAD 180  
 Db 121 NRIELKIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGVSQVLAD 180  
 Qy 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240  
 Db 181 HYQONTPIGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240  
 Qy 241 GSGSGGSG 300  
 Db 239 -----SG 291  
 Qy 301 IGRHKHMFNFDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHKDAVEAFPGGAGMK 360  
 Db 292 IGRHKHMFNFDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHKDAVEAFPGGAGMK 351  
 Qy 361 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWDALFDIVDKONGAITLDENKAY 420  
 Db 352 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWDALFDIVDKONGAITLDENKAY 411  
 Qy 421 TKAAGIIQSSDCEETFRVCDIDESGQLDVDEMTROHLGFWYTMDPACEKLYGGAVP 477  
 Db 412 TKAAGIIQSSDCEETFRVCDIDESGQLDVDEMTROHLGFWYTMDPACEKLYGGAVP 468

RESULT 4  
 ABB08632  
 ID ABB08632 standard; Protein: 450 AA.  
 XX  
 AC ABB08632;  
 XX  
 DT 03-MAY-2002 (first entry)  
 XX  
 DE GFP fusion protein G2A SEQ ID NO 3.  
 XX  
 KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;  
 KW Chemiluminescence Resonance Energy Transfer; CRET;  
 KW central nervous system; neural network.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200192300-A2.  
 XX



Qy 61 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120  
 Db 61 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120  
 Qy 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKOKNGIKANFKIRHNIEDGSVQLAD 180  
 Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKOKNGIKANFKIRHNIEDGSVQLAD 180  
 Qy 181 HYQONTPIGDPVLLPDNHYLSQSALSQKDPNEKRDHMLLEFVTAAGITHGMDELYKSG 240  
 Db 181 HYQONTPIGDPVLLPDNHYLSQSALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK-- 238  
 Qy 241 GSG 300  
 Db 239 -----S GSG 264  
 Qy 301 IGRHKHMFNFDVNHNGKISLDEMVKASDIVINNIGATPEQAKRHKDAVEAFPGGAGMK 360  
 Db 265 IGRHKHMFNFDVNHNGKISLDEMVKASDIVINNIGATPEQAKRHKDAVEAFPGGAGMK 324  
 Qy 361 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIYDKDQNGAITLDEWKAY 420  
 Db 325 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIYDKDQNGAITLDEWKAY 384  
 Qy 421 TKAAGIIQSSDCEETFRVCDIDESGQDLDVDEMTROHLGFWYTMDPACELYGAVP 477  
 Db 385 TKAAGIIQSSDCEETFRVCDIDESGQDLDVDEMTROHLGFWYTMDPACELYGAVP 441

RESULT 6  
 ABB08630  
 ID ABB08630 standard; Protein: 432 AA.  
 XX AC ABB08630;

XX DT 03-MAY-2002 (first entry)

XX DE GFP fusion protein GA SEQ ID NO 1.

XX KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;  
 KW Chemiluminescence Resonance Energy Transfer; CRET;  
 XX central nervous system; neural network.

XX OS Synthetic.

XX FH Key Location/Qualifiers  
 FT Misc-difference 168 /note= "Encoded by CQC"  
 FT Misc-difference 179 /note= "Encoded by OCC"  
 FT Misc-difference 307 /note= "Encoded by GCC"

XX PN WO200192300-A2.

XX PD 06-DEC-2001.

XX XX 01-JUN-2001; 2001WO-EP07057.

XX PR 01-JUN-2000; 2000US-208314P.

XX PR 09-JUN-2000; 2000US-210526P.

XX PR 14-DEC-2000; 2000US-255111P.

XX XX (INSP ) INST PASTEUR.

XX PA (CNRS ) CENT NAT RECH SCI.

XX XX Baubet V, Le Mouellie H, Brulet P;

XX XX WPI; 2002-139605/18.

XX XX N-PSDB; ABA97885.

XX PT New fusion protein comprising a modified bioluminescent system with a

PT fluorescent molecule covalently linked with a photoprotein, useful for  
 PT monitoring calcium fluxes or for detecting electrical activity in a  
 XX group of neural cells -

PS Claim 2; Page 30; 61pp; English.

XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy  
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence  
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule  
 CC covalently linked with a photoprotein. The fusion protein comprises the  
 CC formula: GFP - LINKER - AEQ  
 CC GFP - green fluorescent protein; AEQ - aequorin; and LINKER - a  
 CC polypeptide of 4-63 amino acids.

CC The fusion protein is useful as a bioluminescent Ca<sup>2+</sup> reporter at the  
 CC single cell level. The fusion protein, composition or bioluminescent  
 CC system is useful for monitoring calcium fluxes in real time. This is  
 CC particularly useful for understanding the development, the plasticity and  
 CC the functioning of the central nervous system. The fusion protein  
 CC comprising the bioluminescent system is useful for detecting electrical  
 CC activity in a group of neural cells, for making it possible to complete  
 CC the phenotype study of mutants, or for observing the calcium activity in  
 CC a population of connected cells, for example in a neural network.

XX SQ Sequence 432 AA;

Query Match 89.4%; Score 2287.5; DB 23; Length 432;  
 Best Local Similarity 90.4%; Pred. No. 1.2e-184;  
 Matches 431; Conservative 0; Mismatches 1; Indels 45; Gaps 1;

Qy 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGSDATYGLTKLTKFKTCTCKLPVWPTL 60

Db 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGSDATYGLTKLTKFKTCTCKLPVWPTL 60

Qy 61 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120

Db 61 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120

Qy 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKOKNGIKANFKIRHNIEDGSVQLAD 180

Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKOKNGIKANFKIRHNIEDGSVQLAD 180

Qy 181 HYQONTPIGDPVLLPDNHYLSQSALSQKDPNEKRDHMLLEFVTAAGITHGMDELYKSG 240

Db 181 HYQONTPIGDPVLLPDNHYLSQSALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK-- 238

Qy 241 GSG 300

Db 239 -----SGLRSVKLTSDFNPRW 255

Qy 301 IGRHKHMFNFDVNHNGKISLDEMVKASDIVINNIGATPEQAKRHKDAVEAFPGGAGMK 360

Db 256 IGRHKHMFNFDVNHNGKISLDEMVKASDIVINNIGATPEQAKRHKDAVEAFPGGAGMK 315

Qy 361 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIYDKDQNGAITLDEWKAY 420

Db 316 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIYDKDQNGAITLDEWKAY 375

Qy 421 TKAAGIIQSSDCEETFRVCDIDESGQDLDVDEMTROHLGFWYTMDPACELYGAVP 477

Db 376 TKAAGIIQSSDCEETFRVCDIDESGQDLDVDEMTROHLGFWYTMDPACELYGAVP 432

RESULT 7

ABB76672

ID ABB76672 standard; protein: 655 AA.

XX AC ABB76672;

XX DT 03-SEP-2002 (first entry)

XX DE Protein related to Bombyx mori silk fibroin.

XX KW Silk; fibroin; textile industry.





AAV70779  
ID AAV70779 standard; Protein: 719 AA.  
AC AAV70779;  
DT 31-JUL-2000 (first entry)  
DE EGFP-SMAD2 fusion protein construct.  
KW Fusion construct; EGFP-SMAD2; Green fluorescent protein; cytokine;  
KW cellular response; luminophore; screening; medicament; signal transducer;  
KW light/fluorescence intensity; intracellular signalling pathway; mutant;  
KW fluorescent probe; signal transduction; transforming growth factor-beta;  
KW TGF-beta.  
XX  
XX Chimeric - Homo sapiens.  
OS Chimeric - Aequorea victoria.  
XX  
XX WO200023615-A2.  
PN  
PD 27-APR-2000.  
XX  
XX 15-OCT-1999; 99WO-DK00562.  
XX  
XX 15-OCT-1998; 98DK-0001320.  
XX  
XX (BIOI-) BIOIMAGE AS.  
XX  
XX Arkhammar POG, Terry BR, Scudder KM, Bjorn SP, Thastrup O;  
PI Hagel G;  
PI  
DR WPI: 2000-339710/29.  
DR N-PSDB; RAD00037.  
XX  
XX Quantifying cellular response to an influence, useful for detecting  
PT intracellular translocation or redistribution of biologically active  
PT substances comprises recording changes in spatially distributed light  
PT emitted from a luminophore.  
XX  
XX Example 4; Page 110-112; 150pp; English.  
XX  
XX The patent discloses a method for extracting quantitative information  
CC relating to an influence on a cellular response in mechanically intact  
CC living cells, by recording variation in spatially distributed light  
CC emitted from a luminophore, as a change in light/fluorescence intensity.  
CC The luminophore present in the cells, is capable of being redistributed  
CC in a manner which is related with the degree of the influence. This  
CC method is useful as a screening program, for the identification of a  
CC biologically active substance, that directly or indirectly affects an  
CC intracellular signalling pathway. It is also potentially useful as a  
CC medicament. The fluorescent probe is useful in the back-tracking of  
CC signal transduction pathways. The present sequence is a fusion protein  
CC construct EGFP-SMAD2, comprising the human Smad2 gene, fused to a  
CC derivative of the luminescent green fluorescent protein, EGFP. Smad2 is  
CC a signal transducer, that is induced by some members of the transforming  
CC growth factor-beta (TGF) family of cytokines. This construct is under the  
CC control of a CMV promoter and functions as the luminophore, that is  
CC useful to monitor signalling pathways and identify compounds that  
CC modulate the pathways in living cells.  
XX  
XX Sequence 719 AA;

Query Match 50.68; Score 1295; DB 21; Length 719;  
Best Local Similarity 81.08; Pred. No. 1.5e-100;  
Matches 251; Conservative 9; Mismatches 16; Indels 14; Gaps 1;  
QY 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGDATYGLTKLTKFKICTTGKLPVWPTL 60  
DB 2 VSRGEELFTGVVPIVLVDGVDNGHFKFSVSGEGDATYGLTKLTKFKICTTGKLPVWPTL 61  
QY 61 VTTLTGYVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120  
DB 62 VTTLTGYVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121

QY 121 NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKORNGIKANFKIRHNIEDGSVQLAD 180  
DB 122 NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKORNGIKVNFKIRHNIEDGSVQLAD 181  
QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDNKRDHMLVLEFFVTAAGITTHGMDELYKSG 240  
DB 182 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDNKRDHMLVLEFFVTAAGITTHGMDELYKSG 241  
QY 241 GSG 286  
DB 242 LRSRAQASNSWSSILPTTPVVRKLLGWKKSAGSGSGAGGEGQNGQEKWCEKAVKSIV 301  
QY 287 RSVKLTSDFD 296  
DB 302 KKLKKTGRLD 311

RESULT 10  
ABB08620  
ID ABB08620 standard; Protein: 356 AA.  
XX  
AC ABB08620;  
DT 10-APR-2002 (first entry)  
XX  
DE Yeast polypeptide 1.  
XX  
KW Yeast; drug; cell wall; GPI anchor protein.  
XX  
OS Saccharomyces cerevisiae.  
XX  
PN WO200183733-A1.  
XX  
PD 08-NOV-2001.  
XX  
PF 26-APR-2001; 2001WO-JP03630.  
XX  
PR 01-MAY-2000; 2000JP-0132041.  
XX  
PA (DAUC) DAIICHI PHARM CO LTD.  
XX  
XX Kitamura A, Someya K, Nakajima R;  
XX  
DR WPI: 2002-097496/13.  
DR N-PSDB; ABA97833.  
XX  
XX Screening for drugs that act on cell walls, involves culturing  
PT microorganisms with a reporter protein acting as a GPI anchor protein  
PT in their cell walls -  
XX  
PS Disclosure; Fig 4-5; 44pp; Japanese.

The invention relates to screening for drugs that act on cell walls, a  
comprising culturing microorganisms with a reporter protein acting as a  
GPI anchor protein in their cell walls in the presence of the test  
substance and assaying the amount of reporter protein produced in the  
culture. The present sequence is that of a polypeptide useful to the  
invention.

Sequence 356 AA;  
Query Match 50.58; Score 1293.5; DB 23; Length 356;  
Best Local Similarity 88.08; Pred. No. 7.8e-101;  
Matches 250; Conservative 3; Mismatches 26; Indels 5; Gaps 2;  
QY 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGDATYGLTKLTKFKICTTGKLPVWPTL 60  
DB 40 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGDATYGLTKLTKFKICTTGKLPVWPTL 99  
QY 61 VTTLTGYVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120  
DB 100 VTTLTGYVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 159

QY 121 NRIELKIDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGIKANFKIRHINIEDGSVQLAD 180  
 DB 160 NRIELKIDFKEDGNILGHKLEYNYNHSHVYITADKQKNGIKANFKIRHINIEDGSVQLAD 219  
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDLYK-- 238  
 DB 220 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDLYK-- 279  
 QY 239 --SGSGGG 280  
 DB 280 LGSCTESSPASSNAGSSSKS-NSGSGSGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 322

RESULT 11

AAW85006  
 ID AAW85006 standard; Protein; 631 AA.

AC AAW85006;  
 DT 08-FEB-1999 (first entry)

DE Erkl-green fluorescent protein fusion product.  
 KW Human; Erkl gene; fusion protein; green fluorescent protein; GFP;  
 KW intracellular signalling; chimera.  
 XX Chimeric - Aequorea victoria.  
 OS Chimeric - Homo sapiens.  
 XX W09845704-A2.  
 XX 15-OCT-1998.  
 XX 07-APR-1998; 98WO-DK00145.  
 PR 07-APR-1997; 97DK-0000392.  
 PA (NOVO) NOVO-NORDISK AS.

PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O;  
 PI Tullin S;  
 XX WPI: 1998-594491/50.  
 DR N-PSDB; AAW71022.

XX Determining effect on signalling pathways in live cells from  
 CC redistribution of luminophores - specifically fusions of green  
 CC fluorescent protein with a signalling component, and new apparatus,  
 CC particularly for identifying toxins and potential therapeutic agents  
 PS Example 3; Pages 69-71; 326pp; English.  
 XX The present sequence represents a human Erkl-green fluorescent  
 CC protein fusion product. The fusion protein is used in an assay  
 CC that exemplifies the invention. The specification describes how  
 CC quantitative information about the influence of a molecule on a cellular  
 CC response is obtained by recording the variation, caused by the molecule,  
 CC on mechanically intact living cells, in the spatially distributed light  
 CC emitted from a luminophore present in the cells. The variation in light  
 CC emission is processed to provide information that correlates spatial  
 CC distribution to the degree of the molecule. The method is used to  
 CC identify agents that (indirectly) affect intracellular signalling,  
 CC especially to screen for potential therapeutic agents or toxins, and  
 CC to identify new drug targets.  
 XX Sequence 631 AA;

Query Match 50.5%; Score 1292.5; DB 19; Length 631;  
 Best Local Similarity 58.0%; Pred. No. 2.1e-100;  
 Matches 283; Conservative 27; Mismatches 107; Indels 71; Gaps 11;  
 QY 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYKLTLCFICTTGKLPVWPPTL 60

DB 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGEDATYKLTLCFICTTGKLPVWPPTL 61  
 QY 61 VTTLYGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120  
 DB 62 VTTLYGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 121  
 QY 121 NRIELKIDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGIKANFKIRHINIEDGSVQLAD 180  
 DB 122 NRIELKIDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGIKANFKIRHINIEDGSVQLAD 181  
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDLYKSG 240  
 DB 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDLYKSG 241  
 QY 241 --SGSGGG 280  
 DB 242 LRSRAQASNSTWAAAAAAGGGGGGPRTEGFGVGPCEVEMVKGQDFVGPRTYQIYIG 301  
 QY 281 GQSGGLRS-----VKLTSDFDNPRWIG-----RHKHMFNFDVNHN 316  
 DB 302 EGAYGMVSSAYDHRKTRVAIKKISPFHQYTCQRTLREIQILLRFRHENWIGIRDLRA 361  
 QY 317 GKISLDEMVKASDIVINNLGATPEQAKRHKDAVEAFPG--AGMKYGVETDWPAYIEGW 374  
 DB 362 STLEAMRDVYIVODLMDTDLKLLKSQSLSDHICFYQILRGLK-----YIHS 412  
 QY 375 KKLATD-ELEKYAKNEPTLRIWGDALFIVD--KDONGAIT---LDEW-----K 418  
 DB 413 NVLHRDLKPSNLLSNVTCDLKICDFGLARIADPEHDHGTFLTEYVATRYRAPEIMLSK 472  
 QY 419 AYTKAAGI 426  
 DB 473 GYTKSIDI 480

RESULT 12

AAW70778  
 ID AAW70778 standard; Protein; 631 AA.

AC AAW70778;  
 DT 31-JUL-2000 (first entry)

DE EGFP-Erk1 fusion protein construct.  
 KW Fusion construct; EGFP-Erk1; Green fluorescent protein; luminophore;  
 KW extracellular signal regulated kinase-1; serine/threonine protein kinase;  
 KW mitogen activated protein kinase; MAPK; screening; medicament; mutant;  
 KW cellular response; light/fluorescence intensity; signal transduction;  
 KW intracellular signalling pathway; fluorescent probe.

XX Chimeric - Homo sapiens.  
 OS Chimeric - Aequorea victoria.  
 XX W0200023615-A2.  
 XX 27-APR-2000.  
 XX 15-OCT-1999; 99WO-DK00562.  
 XX 15-OCT-1998; 98DK-0001320.  
 XX (BIOI-) BIOIMAGE AS.  
 PA Arkhammar POG, Terry BR, Scudder KM, Bjorn SP, Thastrup O;  
 PI Hagel G;  
 XX WPI: 2000-339710/29.  
 DR N-PSDB; AAD00036.

PT Quantifying cellular response to an influence, useful for detecting  
 PT intracellular translocation or redistribution of biologically active





PN WO200050872-A2.  
XX 31-AUG-2000.  
FD 25-FEB-2000; 2000WO-US04794.  
PF 26-FEB-1999; 99US-0122152.  
PR 08-MAR-1999; 99US-0123399.  
PR 12-JUL-1999; 99US-0352171.  
XX (CELL-) CELLONICS INC.  
PA Giuliano KA, Kapur R;  
PI WPI; 2000-594086/56.  
XX N-PSDB; AAA93351.  
DR  
DR  
XX Automated cell-based characterization of toxin by contacting cells  
PT containing luminescent reporter molecules with test substance and  
PT analyzing optically  
XX  
XX Example 11; Page 173-174; 336pp; English.  
XX  
XX The invention relates to systems, methods and reagents for cell-based  
CC screening or detection of compounds which affect particular biological  
CC functions. The methods of the invention utilise fluorescent bioreporter  
CC molecules which, when acted on by a compound of interest, cause an  
CC alteration in the cellular distribution of at least the fluorescent  
CC moiety. In one embodiment, the biosensors comprise heat shock proteins  
CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent  
CC protein (GFP), or derivatives thereof). Such biosensors are located in  
CC the cytoplasm, but on stress activation translocate to the nucleus. In  
CC another embodiment bioreporter proteins can be used to detect protease  
CC activity. Such protease bioreporter fusion proteins comprise one or more  
CC fluorescent proteins; a recognition signal which is cleaved by the  
CC protease; and at least one cellular localisation signal. The latter two  
CC components may be components of a single protein which is acted upon by  
CC the protease, or may be from heterologous sources. Due to the  
CC localisation signal, the bioreporter protein is localised to a  
CC particular region of the cell. Once acted on by the protease of interest,  
CC the fluorescent protein is cleaved from the localisation sequence, and  
CC is free to migrate to other locations within the cell. The presence of a  
CC second localisation signal attached to the fluorescent protein enables  
CC the fluorescent protein to be directed to a different cellular  
CC compartment after cleavage of the protease recognition sequence. The  
CC change in distribution of the fluorescent protein can be detected using  
CC imaging methods with a high degree of spatial resolution. The methods  
CC and biosensors of the invention can be used to investigate a wide range  
CC of cellular activities and to screen compounds which modulate these  
CC activities. Biosensors containing a recognition site for caspase, for  
CC example, may be used for the screening of compounds which modulate  
CC apoptosis, while biosensors containing other protease recognition sites  
CC may be used for the detection of proteolytic toxins (such as anthrax  
CC lethal factor). The method provides improved target validation and  
CC candidate compound optimisation by combining many cell screening formats  
CC with fluorescence-based molecular reagents and computer-based feature  
CC extraction, data analysis and automation, resulting in increased  
CC quantity and speed of data collection and faster evaluation of drug  
CC candidates. Sequences AAB22860-B22876 and AAB22936-B22941 represent  
CC biosensor fusion proteins produced in an exemplification of the  
XX invention.  
XX  
XX Sequence 294 AA:  
SQ

Query Match 50.3%; Score 1288.5; DB 21; Length 294;  
Best Local Similarity 91.0%; Pred. No. 1.6e-100;  
Matches 244; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

OY 1 MSKGEELFTGVVILVELDGVNGHKFSVSGEGEDATYGLTKLTKFKICTTKLPVPPPTL 60  
:|||||  
Db 2 VSKGEELFTGVVILVELDGVNGHKFSVSGEGEDATYGLTKLTKFKICTTKLPVPPPTL 61  
:|||||

OY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

Db 62 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121  
OY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180  
Db 122 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLAD 181  
OY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240  
Db 182 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITILGMDELYKSG 241  
OY 241 GSGSGSGSGSGSGG-QSGSGSGSGSG 267  
Db 242 LRSAGAGAGAGAGADEVDGAGADEVDG 269

Search completed: June 17, 2003, 16:40:58  
Job time : 94.3226 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 17:29:13 ; Search time 32.5844 Seconds  
(without alignments)  
430.719 Million cell updates/sec

Title: US-09-863-901-5  
Perfect score: 2560  
Sequence: 1 MSKGEELFTGVVPIVLVDG.....LGFVYWDPAKELKYGAVP 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1290.5	50.4	890	4	US-09-513-783A-174
2	1288.5	50.3	294	4	US-09-513-783A-2
3	1286	50.2	1070	4	US-09-091-042A-2
4	1283	50.1	459	4	US-09-513-783A-170
5	1270	49.6	238	3	US-08-893-327-16
6	1270	49.6	247	3	US-08-893-327-18
7	1269	49.6	783	4	US-09-513-783A-176
8	1267	49.5	941	4	US-09-513-783A-172
9	1266	49.5	1452	4	US-09-127-227-2
10	1265.5	49.4	350	4	US-09-513-783A-30
11	1265	49.4	805	4	US-09-513-783A-178
12	1265	49.4	1407	4	US-08-974-549A-628
13	1263	49.3	302	4	US-09-513-783A-18
14	1262	49.3	238	1	US-08-337-915A-2
15	1262	49.3	238	4	US-09-121-539-1
16	1262	49.3	238	5	PCT-US95-14692-2
17	1259.5	49.2	642	2	US-08-818-253-2
18	1259.5	49.2	642	2	US-08-818-252-2
19	1259.5	49.2	652	2	US-08-818-253-4
20	1259.5	49.2	652	4	US-08-818-252-4
21	1259	49.2	295	4	US-09-513-783A-20
22	1258	49.1	238	1	US-08-753-143-2
23	1258	49.1	238	2	US-08-679-865-2
24	1258	49.1	238	2	US-08-680-876-2
25	1258	49.1	238	2	US-08-792-553-2
26	1258	49.1	238	3	US-08-753-144-2
27	1258	49.1	238	4	US-09-094-359-2

28	1258	49.1	238	4	US-09-172-063-2	Sequence 2, Appl1
29	1258	49.1	238	4	US-09-263-975-2	Sequence 2, Appl1
30	1258	49.1	238	4	US-08-727-452-2	Sequence 2, Appl1
31	1258	49.1	238	4	US-09-418-785-1	Sequence 1, Appl1
32	1257	49.1	292	4	US-09-513-783A-16	Sequence 16, Appl1
33	1257	49.1	347	4	US-09-513-783A-28	Sequence 28, Appl1
34	1256.5	49.1	642	2	US-08-818-253-6	Sequence 6, Appl1
35	1256.5	49.1	642	2	US-08-818-252-6	Sequence 6, Appl1
36	1256.5	49.1	656	2	US-08-818-253-8	Sequence 8, Appl1
37	1256.5	49.1	656	4	US-08-818-252-8	Sequence 8, Appl1
38	1256	49.1	238	3	US-08-643-704A-49	Sequence 49, Appl1
39	1256	49.1	281	4	US-09-062-102-1	Sequence 1, Appl1
40	1256	49.1	281	4	US-09-364-346-1	Sequence 1, Appl1
41	1255	49.0	239	4	US-09-172-063-3	Sequence 3, Appl1
42	1255	49.0	239	4	US-09-513-783A-46	Sequence 46, Appl1
43	1255	49.0	239	4	US-09-316-919-4	Sequence 4, Appl1
44	1255	49.0	282	4	US-09-513-783A-14	Sequence 14, Appl1
45	1255	49.0	296	4	US-09-513-783A-12	Sequence 12, Appl1

## ALIGNMENTS

RESULT 1  
US-09-513-783A-174  
; Sequence 174, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513.783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 174  
; LENGTH: 390  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSC70  
US-09-513-783A-174

Query Match	50.4%	Score 1290.5;	DB 4;	Length 890;		
Best Local Similarity	64.4%	Pred. No. 5e-100;				
Matches	282;	Conservative	22;	Mismatches 77; Indels 57; Gaps 12;		
Qy	1	MSKGEELFTGVVPIVLVDG	VNCHKFSVSGEGDATY	KGKLT	KFKICTTGKLPVWPPTL 60	
Db	2	VSKGEELFTGVVPIVLVDG	VNCHKFSVSGEGDATY	KGKLT	KFKICTTGKLPVWPPTL 61	
Qy	61	VTLYTVGQCFSRYPDHMKQHD	DFKSAPEGYVQERTIFFKDD	CNKT	RAEVKFGDPLV 120	
Db	62	VTLYTVGQCFSRYPDHMKQHD	DFKSAPEGYVQERTIFFKDD	CNKT	RAEVKFGDPLV 121	
Qy	121	NRIELKIDFKEDGNILGHKLE	YNNSHNYIMADKQKNGIKAN	FKIRHN	IEDSGVQLAD 180	
Db	122	NRIELKIDFKEDGNILGHKLE	YNNSHNYIMADKQKNGIKAN	FKIRHN	IEDSGVQLAD 181	
Qy	181	HYQONTPIGDPVLLPDNHYL	STQSALSKDPNERDHMLLE	FTVTA	GITHGMDELKSG 240	
Db	182	HYQONTPIGDPVLLPDNHYL	STQSALSKDPNERDHMLLE	FTVTA	GITHGMDELKSG 241	
Qy	241	-GSGSGSGSGSGSGSGSGSG	SGSGSGSGSGSGSGSGSGSG	SGSGSGSGSGSGSGSGSGSG	SGSGSGSGSGSGSGSGSGSG 283	
Db	242	LRSMSKPAVGIDIGTYSVGV	-FQHGKVEIIANDQGNRTTP	SVAF	TDFERLIGDAK 300	
Qy	284	SGLRSVKLTSDFNPRWIGRH	-----KHMNF	LDVNHNGKIS	IDENYKASDIVI 333	
Db	301	NOVAMNPTNVFADAKRLIGR	FRDDAVVQSDMKH-WPEV	YVNDAGR	KPKV-QVEYK----- 352	
Qy	334	NNLGAT-----PEQAK	-----RHKDA	VEAFG	GAGMKYGVETDNPAYIEGWK	KLATDELEK 384

Db 353 ---GETKFFPEEYSSMLTKKEIAEAYLGKVTNAVTV--PAYENDSORQATKDAGT 407  
QY 385 YAK-----NEPTLIRI 395  
Db 408 IAGLVLRINEPTAAAI 425

## RESULT 2

US-09-513-783A-2  
; Sequence 2, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 294  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: GFP-DEVD-Annexin II construct  
US-09-513-783A-2

Query Match 50.3%; Score 1288.5; DB 4; Length 294;  
Best Local Similarity 91.0%; Pred. No. 1.6e-100;  
Matches 244; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY	1	MSKGEELFGVVPILVELDGDVNGHKFVS	60
Db	2	VSKGEELFGVVPILVELDGDVNGHKFVS	61
QY	61	VTLTYGVQCFSRYPDHMKQHDFFKSAM	120
Db	62	VTLTYGVQCFSRYPDHMKQHDFFKSAM	121
QY	121	NRIELKGIDFKEDGNILGHKLEYN	180
Db	122	NRIELKGIDFKEDGNILGHKLEYN	181
QY	181	HYQONTPIGDGPVLLPDNHYLSTQ	240
Db	182	HYQONTPIGDGPVLLPDNHYLSTQ	241
QY	241	QSGSGGSGSGSGG-QSGSGSGSGSGG	267
Db	242	LRSGAGAGAGAGAGADEVDG	269

## RESULT 3

US-09-091-042A-2  
; Sequence 2, Application US/09091042A  
; Patent No. 6455300  
; GENERAL INFORMATION:

APPLICANT: The Government of the United States of America  
as represented by the Secretary  
Department of Health and Human Services  
Washington, D.C.  
Hun Ph.D., Han  
Hager Ph.D., Gordon L.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MONITORING  
DNA BINDING MOLECULES IN LIVING CELLS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSER: Needle & Rosenberg  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta

STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/091,042A  
FILING DATE: 08-Jun-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/008,373  
FILING DATE: 08 Dec 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Selby, Elizabeth  
REGISTRATION NUMBER: 38298  
REFERENCE/DOCKET NUMBER: 14014.0183  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-688-0770  
TELEFAX: 404-688-9880  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1070 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-091-042A-2

Query Match	50.28;	Score	1286;	DB	4;	Length	1070;
Best Local Similarity	97.68;	Pred. No.	1.6e-99;				
Matches	239;	Conservative	2;	Mismatches	4;	Indels	0;
Gaps	0;						
QY	1	MSKGEELFGVVPILVELDGDVNGHKFVS	60				
Db	23	MSKGEELFGVVPILVELDGDVNGHKFVS	82				
QY	61	VTLTYGVQCFSRYPDHMKQHDFFKSAM	120				
Db	83	VTLTYGVQCFSRYPDHMKQHDFFKSAM	142				
QY	121	NRIELKGIDFKEDGNILGHKLEYN	180				
Db	143	NRIELKGIDFKEDGNILGHKLEYN	202				
QY	181	HYQONTPIGDGPVLLPDNHYLSTQ	240				
Db	203	HYQONTPIGDGPVLLPDNHYLSTQ	262				
QY	241	GSGSG 245					
Db	263	GAGAG 267					

## RESULT 4

US-09-513-783A-170  
; Sequence 170, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 170  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Artificial Sequence





```

RESULT 7
US-09-513-783A-176
; Sequence 176, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-11
; CURRENT APPLICATION NUMBER: US/09/513, 783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 176
; LENGTH: 783
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSFL
US-09-513-783A-176

```

Query Match	49.6%	Score 1269;	DB 4;	Length 783;
Best Local Similarity	91.0%;	Pred. No. 2.7e-98;		
Matches 242;	Conservative 3;	Mismatches 21;	Indels 0;	Gaps 0;

  

QY	1	MSKGEELFTGVVPI	LVELDGVNGHKF	SVSGEGDATY	GKLT	LFICIT	GKLPVPPPTL	60
		:	:	:	:	:	:	:
Db	2	VSKGEELFTGVVPI	LVELDGVNGHKF	SVSGEGDATY	GKLT	LFICIT	GKLPVPPPTL	61
		:	:	:	:	:	:	:
QY	61	VTTLTYGVQCF	SRYPDMKHQD	FFKSAMPEGV	QVQERT	IFFKDDGNY	KTRAEVKEGDTLV	120
		:	:	:	:	:	:	:
Db	62	VTTLTYGVQCF	SRYPDMKHQD	FFKSAMPEGV	QVQERT	IFFKDDGNY	KTRAEVKEGDTLV	121
		:	:	:	:	:	:	:
QY	121	NRIELKGI	DFKEDGNILGH	KLLEYN	NSHNVI	IMADQKNG	IKANFKITHNIEDGSVOLAD	180
		:	:	:	:	:	:	:
Db	122	NRIELKGI	DFKEDGNILGH	KLLEYN	NSHNVI	IMADQKNG	IKVNFIIKHNIEDGSVOLAD	181
		:	:	:	:	:	:	:
QY	181	HYQNTPTIGDGP	VLIPDNH	YLSLTSQ	SALSKD	PNPEKRDH	VMVLEFVTAAGITHGMDELYKSG	240
		:	:	:	:	:	:	:
Db	182	HYQNTPTIGDGP	VLIPDNH	YLSLTSQ	SALSKD	PNPEKRDH	VMVLEFVTAAGITILGMDELYKSG	241
		:	:	:	:	:	:	:
QY	241	GGSGGGQSGG	SGSGGGSGGGGS	266				
		:	:	:	:	:	:	:
Db	242	LRSSRAQASNS	AVEMDPLPVG	FGAAGPS	267			
		:	:	:	:	:	:	:

RESULT 8  
US-09-513-783A-172  
; Sequence 172, Application US/09513783A  
; Patent No. 6416959

```

; GENERAL INFORMATION:
; APPLICANT:  Giuliano, Kenneth A.
; APPLICANT:  Kapur, Ravi
; TITLE OF INVENTION:  A System for Cell Based Screening
; FILE REFERENCE:  97-022-L1
; CURRENT APPLICATION NUMBER:  US/09/513,783A
; CURRENT FILING DATE:  2000-02-25
; NUMBER OF SEQ ID NOS:  180
; SOFTWARE:  PatentIn ver. 2.0
; SEQ ID NO 172
; LENGTH:  941
; TYPE:  prt
; ORGANISM:  Artificial Sequence
; FEATURE:
; OTHER INFORMATION:  Description of Artificial Sequence:  GFP-HSP70
US-09-513-783A-172

```

[illegible]

```

RESULT 9
US-09-127-227-2
; Sequence 2, Application US/09127227
; Patent No. 6399354
; GENERAL INFORMATION:
; APPLICANT: David M. Koipe
; APPLICANT: Travis J. Taylor
; APPLICANT: Elizabeth McNamee
; TITLE OF INVENTION: Replication-Competent Virus Expressing A
; TITLE OF INVENTION: Fusion Protein
; FILE REFERENCE: H098-05
; CURRENT APPLICATION NUMBER: US/09/127,227
; CURRENT FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1452
; TYPE: PRT
; ORGANISM: herpesvirus
US-09-127-227-2

```

Query Match	49.5%;	Score 1266;	DB 4;	Length 1452;
Best Local Similarity	99.2%;	Pred. No. 1.1e-97;		
Matches 236;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

  

QY	1	MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLTKLKICTTGKLPVDPWPTL	60
Db	1215	MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLTKLKICTTGKLPVDPWPTL	1274
QY	61	VTTLYIGVQCFSRYPDHMKQHDFFSKAMPEGVYQERTIFFKDDGNTKRAEVKFGEDTLV	120

Db 1275 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 1334  
Qy 121 NRIELKGIDFKEDGNILGHKLLEYNHNHNYIMADKQKNGIKANFKIRHNTEDEGSVOLAD 180  
Db 1335 NRIELKGIDFKEDGNILGHKLLEYNHNHNYIMADKQKNGIKANFKIRHNTEDEGSVOLAD 1394  
Qy 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITHGMDELK 238  
Db 1395 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITHGMDELK 1452

RESULT 10  
US-09-513-783A-30  
; Sequence 30, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 30  
; LENGTH: 350  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: NLS-Fred25-cellubrevin construct  
US-09-513-783A-30

Query Match 49.4%; Score 1265.5; DB 4; Length 350;  
Best Local Similarity 93.4%; Pred. No. 1.7e-98;  
Matches 239; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Qy 2 SKGEELFTGVVPILVELDGVNKGHSVSGEGDATYGKLTFLKFICTTGKLPVPMPTLV 61  
Db 9 SKGEELFTGVVPILVELDGVNKGHSVSGEGDATYGKLTFLKFICTTGKLPVPMPTLV 68  
Qy 62 TLTLYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLVN 121  
Db 69 TLTLYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLVN 128  
Qy 122 NRIELKGIDFKEDGNILGHKLLEYNHNHNYIMADKQKNGIKANFKIRHNTEDEGSVOLADH 181  
Db 129 NRIELKGIDFKEDGNILGHKLLEYNHNHNYIMADKQKNGIKANFKIRHNTEDEGSVOLADH 188  
Qy 182 YQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITHGMDELKSGG 241  
Db 189 YQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITHGMDELKSGG 248  
Qy 242 SGSGGSGSGSGSGSGGS 257  
Db 249 S-TGVPSGSSAATGSN 263

RESULT 11  
US-09-513-783A-178  
; Sequence 178, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.  
; APPLICANT: Kapur, Ravi  
; TITLE OF INVENTION: A System for Cell Based Screening  
; FILE REFERENCE: 97-022-L1  
; CURRENT APPLICATION NUMBER: US/09/513,783A  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 178  
; LENGTH: 805

; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GFP-NFKB  
US-09-513-783A-178

Query Match 49.4%; Score 1265; DB 4; Length 805;  
Best Local Similarity 98.8%; Pred. No. 6.1e-98;  
Matches 237; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVVPILVELDGVNKGHSVSGEGDATYGKLTFLKFICTTGKLPVPMPTLV 60  
Db 2 VSKGEELFTGVVPILVELDGVNKGHSVSGEGDATYGKLTFLKFICTTGKLPVPMPTLV 61  
Qy 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 120  
Db 62 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 121  
Qy 121 NRIELKGIDFKEDGNILGHKLLEYNHNHNYIMADKQKNGIKANFKIRHNTEDEGSVOLAD 180  
Db 122 NRIELKGIDFKEDGNILGHKLLEYNHNHNYIMADKQKNGIKANFKIRHNTEDEGSVOLAD 181  
Qy 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITHGMDELKSG 240  
Db 182 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITHGMDELKSG 241

RESULT 12  
US-09-974-549A-628  
; Sequence 628, Application US/08974549A  
; Patent No. 6166178  
; GENERAL INFORMATION:  
; APPLICANT: Cech, Thomas R.  
; APPLICANT: Lingner, Joachim  
; APPLICANT: Nakamura, Toru  
; APPLICANT: Chapman, Karen B.  
; APPLICANT: Morin, Gregg B.  
; APPLICANT: Harley, Calvin B.  
; APPLICANT: Andrews, William H.  
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
; NUMBER OF SEQUENCES: 727  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/974,549A  
; FILING DATE: 19-NOV-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/724,643  
; FILING DATE: 01-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/844,419  
; FILING DATE: 18-APR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/846,017  
; FILING DATE: 25-APR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/851,843  
; FILING DATE: 06-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/854,050  
; FILING DATE: 09-MAY-1997  
; PRIOR APPLICATION DATA:

Query Match 49.4%; Score 1265; DB 4; Length 1407;  
Best Local Similarity 98.8%; Pred. No. 1.3e-97;  
Matches 237; Conservative 1; Mismatches 2.

APPLICANT: Helm, Roger  
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Robbins, Berliner & Carson  
STREET: 201 No. 5625048th Figueroa Street, Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90012  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/337,915A  
FILING DATE: 08/29/90

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Spitals, John P.  
REGISTRATION NUMBER: 29,215  
REFERENCE/DOCKET NUMBER: 1379-178  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 977-1001  
TELEFAX: (213) 977-1003  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
-337-915A-2

US-08-337-915A-2

Query Match 49.3%; Score 1262; DB 1; Length 238;  
 Best Local Similarity 98.7%; Pred. No. 2e-98;  
 Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVVELDGVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVPWPTL 60  
 DB 1 MSKGEELFTGVVPIVVELDGVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVPWPTL 60

QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVFEGETLV 120  
 DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVFEGETLV 120

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIINADKOKNGIKANFKIRHNIEDGSVOLAD 180  
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIINADKOKNGIKANFKIRHNIEDGSVOLAD 180

QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDLYK 238  
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDLYK 238

RESULT 15

US-09-121-539-1  
 ; Sequence 1, Application US/09121539B  
 ; Patent No. 6194548  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Osumi, Takashi  
 ; APPLICANT: Tsukamoto, Toshiro  
 ; APPLICANT: Tsukamoto, No. 61945481yo  
 ; APPLICANT: Yamasaki, Masatoshi  
 ; TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT  
 ; TITLE OF INVENTION: PROTEINS  
 ; FILE REFERENCE: 046124-5005  
 ; CURRENT APPLICATION NUMBER: US/09/121,539B  
 ; CURRENT FILING DATE: 1998-07-24  
 ; PRIOR APPLICATION NUMBER: JP 026418/1998  
 ; PRIOR FILING DATE: 1998-01-23  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 238  
 ; TYPE: PRT  
 ; ORGANISM: Aequorea victoria  
 ; FEATURE:  
 ; OTHER INFORMATION: Green fluorescent protein  
 US-09-121-539-1

Query Match 49.3%; Score 1262; DB 4; Length 238;  
 Best Local Similarity 98.7%; Pred. No. 2e-98;  
 Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVVELDGVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVPWPTL 60  
 DB 1 MSKGEELFTGVVPIVVELDGVNGHKFVSVEGEGDATYGLTKLKFICTTGKLPVPWPTL 60

QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVFEGETLV 120  
 DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVFEGETLV 120

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIINADKOKNGIKANFKIRHNIEDGSVOLAD 180  
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIINADKOKNGIKANFKIRHNIEDGSVOLAD 180

QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDLYK 238  
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDLYK 238

Search completed: June 18, 2003, 17:32:55  
 Job time : 34.5844 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 17:31:48 ; Search time 57.9279 Seconds  
(without alignments)  
891.015 Million cell updates/sec

Title: US-09-863-901-5  
Perfect score: 2560  
Sequence: 1 MSKGEELFTGVVPIVLVDG.....LGFWTMDPACRKLKYGAVP 477

Scoring table: BLOSUM62

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications\_AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1295	50.6	719	9	US-10-072-036-51
2	1292.5	50.5	631	9	US-10-072-036-39
3	1290.5	50.4	890	9	US-10-100-957A-174
4	1288.5	50.3	294	9	US-10-100-957A-2
5	1286	50.2	1070	9	US-10-001-486B-2
6	1286	50.2	1099	9	US-10-259-864-4
7	1286	50.2	1147	9	US-10-259-864-1
8	1283.5	50.1	633	9	US-10-072-036-45
9	1283	50.1	459	9	US-10-100-957A-170
10	1278.5	49.9	635	9	US-10-072-036-125
11	1276.5	49.9	1171	9	US-10-072-036-131
12	1271	49.6	397	9	US-10-033-717-30
13	1271	49.6	429	9	US-10-033-717-29
14	1269.5	49.6	359	9	US-10-033-717-33
15	1269.5	49.6	359	9	US-10-033-717-34
16	1269.5	49.6	391	9	US-10-033-717-32
17	1269.5	49.6	997	9	US-10-072-036-121
18	1269	49.6	607	9	US-10-072-036-47
19	1269	49.6	783	9	US-10-100-957A-176

20	1268	49.5	403	9	US-10-033-717-31	Sequence 31, Appl
21	1268	49.5	842	9	US-10-072-036-43	Sequence 43, Appl
22	1267	49.5	605	9	US-10-072-036-41	Sequence 41, Appl
23	1267	49.5	727	9	US-10-072-036-139	Sequence 139, App
24	1267	49.5	941	9	US-10-100-957A-172	Sequence 172, App
25	1266.5	49.5	806	9	US-10-072-036-53	Sequence 53, Appl
26	1266	49.5	933	9	US-10-072-036-135	Sequence 135, App
27	1266	49.5	1089	9	US-10-259-864-2	Sequence 2, Appl1
28	1266	49.5	1452	12	US-10-050-673-2	Sequence 2, Appl1
29	1265.5	49.4	350	9	US-10-100-957A-30	Sequence 30, Appl1
30	1265	49.4	308	9	US-10-033-717-35	Sequence 35, Appl
31	1265	49.4	442	9	US-10-072-036-127	Sequence 127, App
32	1265	49.4	544	9	US-10-072-036-115	Sequence 115, App
33	1265	49.4	797	9	US-10-072-036-143	Sequence 143, App
34	1265	49.4	805	9	US-10-100-957A-178	Sequence 178, App
35	1265	49.4	843	9	US-10-072-036-117	Sequence 117, App
36	1265	49.4	871	9	US-10-072-036-109	Sequence 109, App
37	1265	49.4	968	9	US-10-072-036-49	Sequence 49, Appl
38	1265	49.4	1039	9	US-10-072-036-55	Sequence 55, Appl
39	1265	49.4	1090	9	US-10-259-864-6	Sequence 6, Appl1
40	1265	49.4	1407	9	US-10-044-592-334	Sequence 334, App
41	1265	49.4	1407	9	US-10-044-539-334	Sequence 334, App
42	1263	49.3	302	9	US-10-100-957A-18	Sequence 18, Appl
43	1262	49.3	238	9	US-09-866-538-2	Sequence 2, Appl1
44	1262	49.3	238	9	US-09-900-345A-125	Sequence 125, App
45	1262	49.3	238	9	US-10-121-258-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1  
US-10-072-036-51  
; Sequence 51, Application US/10072036  
; Publication No. US20030082564A1  
; GENERAL INFORMATION:  
; APPLICANT: Ole THASTRUP  
; APPLICANT: Sara BJORN  
; APPLICANT: Soren TULLIN  
; APPLICANT: Kasper ALMHOLT  
; APPLICANT: Kurt SCUDDER  
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To A  
; FILE REFERENCE: 3759-0120P  
; CURRENT APPLICATION NUMBER: US/10/072,036  
; PRIOR FILING DATE: 2002-09-13  
; PRIOR APPLICATION NUMBER: 09/417,197  
; PRIOR FILING DATE: 1999-10-07  
; NUMBER OF SEQ ID NOS: 143  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 51  
; LENGTH: 719  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: EGFP-Smad2 fusion  
US-10-072-036-51

Query Match	50.6%	Score 1295;	DB 9;	Length 719;
Best Local Similarity	81.0%;	Pred. No. 7.9e-82;		
Matches 251;	Conservative	9;	Mismatches 36;	Indels 14; Gaps 1;
Qy	1	MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYGKLTLLKFTCTTGKLPVWPPTL	60	
Db	2	VSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDGYGKLTLLKFTCTTGKLPVWPPTL	61	
Qy	61	VTILTYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERIFPFDDGNYKTRAEVKEGDTLV	120	
Db	62	VTTLYGVQCFSRYPDHMKQHDFFKFSAMPEGYVQERIFPFDDGNYKTRAEVKEGDTLV	121	
Qy	121	NRIELKGIDGKFDGNIILGHKLEYNINSHNYYIMADKOKNGIKANFKIRHNIEDGSVQLAD	180	
Db	122	NRIELKGIDGKFDGNIILGHKLEYNINSHNYYIMADKOKNGIKANFKIRHNIEDGSVQLAD	181	

Query Match	50.5%	Score 1292.5;	DB 9;	Length 631;
Best Local Similarity	58.0%;	Pred. No. 1e-81;		
Matches 283;	Conservative 27;	Mismatches 107;	Indels 71;	Gaps 11;
1 MSKGEELFTGVVPLVELGDVNGHKFVS	GGEGDATY	GKLT	FKICTT	GKLPVMPPTL 60
2 VSKGEELFTGVVPLVELGDVNGHKFVS	GGEGDATY	GKLT	FKICTT	GKLPVMPPTL 61
61 VTTLTGYQCFSRYPDHMKQHDFFKSAM	PEGVYQERTIFF	KDDGNYKTR	ARVKEG	TGLV 120
62 VTTLTGYQCFSRYPDHMKQHDFFKSAM	PEGVYQERTIFF	KDDGNYKTR	ARVKEG	TGLV 121
121 NRTELKGDIDKDGNIHLGKLEYNVNSH	VNYIMADKQNGIKANFKIRHN	IEDSGSVQ	LAD 180	
122 NRTELKGDIDKDGNIHLGKLEYNVNSH	VNYIMADKQNGIKANFKIRHN	IEDSGSVQ	LAD 181	
181 HYQONTPIGDGPVLLPDNHYLSTQSAL	KDNPNEKRDHMLLEFVTAAGITG	MDLYKSG 240		
182 HYQONTPIGDGPVLLPDNHYLSTQSAL	KDNPNEKRDHMLLEFVTAAGITG	MDLYKSG 241		
241 --GSGSGGSGSGSGSGSGSGSGSGSG	-----CQSGSGG	-----S 280		
242 LRSRQAQSNSTWAAAAAQGGGGEPR	TGEGVPGVPEVMVKQPPDVGPRT	YQLQYIG 301		
281 GGOSGLRS	-----VKLTSDFDONPWIG	-----RHKMFENFLDVNHN 316		
302 EGAYGMVSSAYDHRVKTTRVAIKKIS	PFPHQYQCORTLREIQILLRF	RHEYNIGIRLLRA 361		
317 GKISLDEWYKASDIVINNLGATPQAKR	HKDAVEAFEGG--AGMKYGVETD	WPAYIEGW 374		
362 STLEAMRDYIVODLWETDLYKLKLSQ	LSNDHCYFLYQLLRGLK-----YI	HA 412		

RESULT 4  
US-10-100-957A-2  
; Sequence 2, Application US/10100957A  
; Publication No. US20030096322A1  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.





```

Db 61 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRIELKGIDFEDKNIGLKLEYNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGIDFEDKNIGLKLEYNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
Db 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKGA 240
QY 241 GSGSG 245
Db 241 GAGAG 245

```

## RESULT 7

```

US-10-259-864-1
; Sequence 1, Application US/10259864
; Publication No. US20030077645A1
; GENERAL INFORMATION:
; APPLICANT: Hager, Gordon L
; TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfamily
; TITLE OF INVENTION: Receptor Ligands, and Methods and Kits For Detecting and Charact
; TITLE OF INVENTION: Ligands
; FILE REFERENCE: 24799
; CURRENT APPLICATION NUMBER: US/10/259,864
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/325,178
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Rat/human chimera
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (1)..()
; OTHER INFORMATION: Chimeric Protein
US-10-259-864-1

```

```

Query Match 50.2%; Score 1286; DB 9; Length 1147;
Best Local Similarity 97.6%; Pred. No. 5.7e-81;
Matches 239; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGKLTIKFTCTTGKLPVWPPTL 60
Db 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGKLTIKFTCTTGKLPVWPPTL 60
QY 61 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 61 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRIELKGIDFEDKNIGLKLEYNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGIDFEDKNIGLKLEYNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
Db 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKGA 240
QY 241 GSGSG 245
Db 241 GAGAG 245

```

## RESULT 8

```

US-10-072-036-45
; Sequence 45, Application US/10072036
; Publication No. US20030082564A1
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP

```

```

; APPLICANT: Sara BJRON
; APPLICANT: Soren TULLIN
; APPLICANT: Kasper ALMHOLT
; APPLICANT: Kurt SCUDDER
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To A
; FILE REFERENCE: 3759-0120P
; CURRENT APPLICATION NUMBER: US/10/072,036
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/417,197
; PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 45
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-Jnk1 fusion
US-10-072-036-45

```

```

Query Match 50.1%; Score 1283.5; DB 9; Length 633;
Best Local Similarity 49.6%; Pred. No. 4.3e-81;
Matches 298; Conservative 40; Mismatches 104; Indels 159; Gaps 17;
QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGKLTIKFTCTTGKLPVWPPTL 60
Db 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGKLTIKFTCTTGKLPVWPPTL 61
QY 61 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 62 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
QY 121 NRIELKGIDFEDKNIGLKLEYNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 122 NRIELKGIDFEDKNIGLKLEYNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 181
QY 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
Db 182 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK-- 239
QY 241 GSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 286
Db 240 --SGLRSARAIMSRKRDNNFYSGEIGDSTFTVLKRYQNLKPIGSGAQGIVCAAYDAILE 298
QY 287 RSV---KLTSDFDNPRWIGR-----HKMFNFDLVNHNKISLDEM--VY---- 326
Db 299 RNAIKLSRPFQNTAKRAYRELVMKCVNHNKIIGLLNVFTPOK--SLEEFQDVYIVM 357
QY 327 -----KASDIVI----- 333
Db 358 ELM DANLCQVIQMELDHERMSYLLYQMLCGIKHLHSAGIHRDLKPSNIVKSDCTLKL 417
QY 334 -----NLLGATPEQAKRHKDAVEAFGGAGMKYGVETDWA----- 369
Db 418 DFGLARTAGTSPMTYVTVRYRAPEVIL--GMYKENVDL--WSVCGINGEMVCHKILFP 475
QY 370 ----YIEGWKKLATD-----ELEYAKNEP-----TLIRWGDAFDIVDKD 407
Db 476 GRDYIDQWKNVIEQLGTPCPEFMKKLQPTVRYVENRPFYAGYSFEKLPDVLFP--ADSE 534
QY 408 QNGAITLDWKAYTAAGTIQSSECEETFRVCDIDESQLDVDE--MTRQHLGFWYMDP 466
Db 535 HNK-----LKASQARDLLSKMLVIDASKRISVDEALQHPYINWY--DP 576
QY 467 A 467
Db 577 S 577

```

## RESULT 9

```

US-10-100-957A-170
; Sequence 170, Application US/10100957A

```

Qy	61	VTTLTTVGVCFSRYPDHMKQHDFKSA MPEGYVQERTIFFKDDGNKYTRAEVAFEGDTLV	120
Db	62	VTTLTTVGVCFSRYPDHMKQHDFKSA MPEGYVQERTIFFKDDGNKYTRAEVAFEGDTLV	121

```
US-10-072-036-131
Query Match          49.9%; Score 1276.5; DB 9; Length 1171;
Best Local Similarity 90.7%; Pred. No. 2.7e-80;
Matches 244; Conservative 3; Mismatches 21; Indels 1; Gaps 1;

QY   1 MSKGEELFTGVVPIVELDGDVNGHKFVSVEGEGDATYKLTILKFICTTGTGLPVPPTL 60
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:
Db    2 VSKGEELFTGVVPIVELDGDVNGHKFVSVEGEGDATYKLTILKFICTTGTGLPVPPTL 61
```



; APPLICANT: TOPOI, LILIA  
; APPLICANT: MARX, MARIA  
; APPLICANT: CALOTHY, GEORGES  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DRM, A SECRETED PROTEIN  
; TITLE OF INVENTION: WITH CELL GROWTH INHIBITING ACTIVITY  
; FILE REFERENCE: 14014.0358  
; CURRENT APPLICATION NUMBER: US/10/033,717  
; CURRENT FILING DATE: 2001-12-27  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/444,066  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-19  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/277,407  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/079,440  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-26  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 33  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: /No. US20030078406ale -  
; OTHER INFORMATION: synthetic construct  
US-10-033-717-33

Query Match 49.6%; Score 1269.5; DB 9; Length 359;  
Best Local Similarity 89.0%; Pred. No. 2e-80;  
Matches 242; Conservative 4; Mismatches 7; Indels 19; Gaps 1;  
QY 1 MSKGEELFTGVVPIVVELDGDVNGHFKFSVSGEGEGDATYKGLTKFKICTTGKLPVPWPTL 60  
Db 2 VSKGEELFTGVVPIVVELDGDVNGHFKFSVSGEGEGDATYKGLTKFKICTTGKLPVPWPTL 61  
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120  
Db 62 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121  
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180  
Db 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKVNFKIRHNIEDGSVQLAD 181  
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNPKRDHMLLEFVTAAGITLGMDELYKSG 240  
Db 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNPKRDHMLLEFVTAAGITLGMDELYKSG 241  
QY 241 -----GSGSGGSGSGSGS 253  
Db 242 LRMRAQHNDSEQTQSPPPQPSRTRGRGGRGT 273

RESULT 15  
US-10-033-717-34  
; Sequence 34, Application US/10033717  
; Publication No. US20030078406A1  
; GENERAL INFORMATION:  
; APPLICANT: BLAIR, DONALD  
; APPLICANT: CLAUSEN, PETER  
; APPLICANT: TOPOI, LILIA  
; APPLICANT: MARX, MARIA  
; APPLICANT: CALOTHY, GEORGES  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DRM, A SECRETED PROTEIN  
; TITLE OF INVENTION: WITH CELL GROWTH INHIBITING ACTIVITY  
; FILE REFERENCE: 14014.0358  
; CURRENT APPLICATION NUMBER: US/10/033,717  
; CURRENT FILING DATE: 2001-12-27  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/444,066  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-19  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/277,407  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/079,440  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-26  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 34  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: /No. US20030078406ale -  
; OTHER INFORMATION: synthetic construct  
US-10-033-717-34

Query Match 49.6%; Score 1269.5; DB 9; Length 359;  
Best Local Similarity 89.0%; Pred. No. 2e-80;  
Matches 242; Conservative 4; Mismatches 7; Indels 19; Gaps 1;  
QY 1 MSKGEELFTGVVPIVVELDGDVNGHFKFSVSGEGEGDATYKGLTKFKICTTGKLPVPWPTL 60  
Db 2 VSKGEELFTGVVPIVVELDGDVNGHFKFSVSGEGEGDATYKGLTKFKICTTGKLPVPWPTL 61  
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120  
Db 62 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121  
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180  
Db 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKVNFKIRHNIEDGSVQLAD 181  
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNPKRDHMLLEFVTAAGITLGMDELYKSG 240  
Db 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNPKRDHMLLEFVTAAGITLGMDELYKSG 241  
QY 241 -----GSGSGGSGSGSGS 253  
Db 242 LRMRAQHNDSEQTQSPPPQPSRTRGRGGRGT 273

Search completed: June 18, 2003, 17:44:06  
Job time : 58.9279 secs



A: Accession: 030301  
A: Molecule type: mRNA  
A: Residues: 1-24, 'Q', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208, 'R', 210-211, 'K', 213-214, 'R', 216-217, 'K', 219-220, 'R', 222-223, 'K', 225-226, 'R', 228-229, 'K', 231-232, 'R', 234-235, 'K', 237-238, 'R', 240-241, 'K', 243-244, 'R', 246-247, 'K', 249-250, 'R', 252-253, 'K', 255-256, 'R', 258-259, 'K', 261-262, 'R', 264-265, 'K', 267-268, 'R', 270-271, 'K', 273-274, 'R', 276-277, 'K', 279-280, 'R', 282-283, 'K', 285-286, 'R', 288-289, 'K', 291-292, 'R', 294-295, 'K', 297-298, 'R', 300-301, 'K', 303-304, 'R', 306-307, 'K', 309-310, 'R', 312-313, 'K', 315-316, 'R', 318-319, 'K', 321-322, 'R', 324-325, 'K', 327-328, 'R', 330-331, 'K', 333-334, 'R', 336-337, 'K', 339-340, 'R', 342-343, 'K', 345-346, 'R', 348-349, 'K', 351-352, 'R', 354-355, 'K', 357-358, 'R', 360-361, 'K', 363-364, 'R', 366-367, 'K', 369-370, 'R', 372-373, 'K', 375-376, 'R', 378-379, 'K', 381-382, 'R', 384-385, 'K', 387-388, 'R', 390-391, 'K', 393-394, 'R', 396-397, 'K', 399-400, 'R', 402-403, 'K', 405-406, 'R', 408-409, 'K', 411-412, 'R', 414-415, 'K', 417-418, 'R', 420-421, 'K', 423-424, 'R', 426-427, 'K', 429-430, 'R', 432-433, 'K', 435-436, 'R', 438-439, 'K', 441-442, 'R', 444-445, 'K', 447-448, 'R', 450-451, 'K', 453-454, 'R', 456-457, 'K', 459-460, 'R', 462-463, 'K', 465-466, 'R', 468-469, 'K', 471-472, 'R', 474-475, 'K', 477-478, 'R', 480-481, 'K', 483-484, 'R', 486-487, 'K', 489-490, 'R', 492-493, 'K', 495-496, 'R', 498-499, 'K', 501-502, 'R', 504-505, 'K', 507-508, 'R', 510-511, 'K', 513-514, 'R', 516-517, 'K', 519-520, 'R', 522-523, 'K', 525-526, 'R', 528-529, 'K', 531-532, 'R', 534-535, 'K', 537-538, 'R', 540-541, 'K', 543-544, 'R', 546-547, 'K', 549-550, 'R', 552-553, 'K', 555-556, 'R', 558-559, 'K', 561-562, 'R', 564-565, 'K', 567-568, 'R', 570-571, 'K', 573-574, 'R', 576-577, 'K', 579-580, 'R', 582-583, 'K', 585-586, 'R', 588-589, 'K', 591-592, 'R', 594-595, 'K', 597-598, 'R', 600-601, 'K', 603-604, 'R', 606-607, 'K', 609-610, 'R', 612-613, 'K', 615-616, 'R', 618-619, 'K', 621-622, 'R', 624-625, 'K', 627-628, 'R', 630-631, 'K', 633-634, 'R', 636-637, 'K', 639-640, 'R', 642-643, 'K', 645-646, 'R', 648-649, 'K', 651-652, 'R', 654-655, 'K', 657-658, 'R', 660-661, 'K', 663-664, 'R', 666-667, 'K', 669-670, 'R', 672-673, 'K', 675-676, 'R', 678-679, 'K', 681-682, 'R', 684-685, 'K', 687-688, 'R', 690-691, 'K', 693-694, 'R', 696-697, 'K', 699-700, 'R', 702-703, 'K', 705-706, 'R', 708-709, 'K', 711-712, 'R', 714-715, 'K', 717-718, 'R', 720-721, 'K', 723-724, 'R', 726-727, 'K', 729-730, 'R', 732-733, 'K', 735-736, 'R', 738-739, 'K', 741-742, 'R', 744-745, 'K', 747-748, 'R', 750-751, 'K', 753-754, 'R', 756-757, 'K', 759-760, 'R', 762-763, 'K', 765-766, 'R', 768-769, 'K', 771-772, 'R', 774-775, 'K', 777-778, 'R', 780-781, 'K', 783-784, 'R', 786-787, 'K', 789-790, 'R', 792-793, 'K', 795-796, 'R', 798-799, 'K', 801-802, 'R', 804-805, 'K', 807-808, 'R', 810-811, 'K', 813-814, 'R', 816-817, 'K', 819-820, 'R', 822-823, 'K', 825-826, 'R', 828-829, 'K', 831-832, 'R', 834-835, 'K', 837-838, 'R', 840-841, 'K', 843-844, 'R', 846-847, 'K', 849-850, 'R', 852-853, 'K', 855-856, 'R', 858-859, 'K', 861-862, 'R', 864-865, 'K', 867-868, 'R', 870-871, 'K', 873-874, 'R', 876-877, 'K', 879-880, 'R', 882-883, 'K', 885-886, 'R', 888-889, 'K', 891-892, 'R', 894-895, 'K', 897-898, 'R', 900-901, 'K', 903-904, 'R', 906-907, 'K', 909-910, 'R', 912-913, 'K', 915-916, 'R', 918-919, 'K', 921-922, 'R', 924-925, 'K', 927-928, 'R', 930-931, 'K', 933-934, 'R', 936-937, 'K', 939-940, 'R', 942-943, 'K', 945-946, 'R', 948-949, 'K', 951-952, 'R', 954-955, 'K', 957-958, 'R', 960-961, 'K', 963-964, 'R', 966-967, 'K', 969-970, 'R', 972-973, 'K', 975-976, 'R', 978-979, 'K', 981-982, 'R', 984-985, 'K', 987-988, 'R', 990-991, 'K', 993-994, 'R', 996-997, 'K', 999-1000, 'R', 1002-1003, 'K', 1005-1006, 'R', 1008-1009, 'K', 1011-1012, 'R', 1014-1015, 'K', 1017-1018, 'R', 1020-1021, 'K', 1023-1024, 'R', 1026-1027, 'K', 1029-1030, 'R', 1032-1033, 'K', 1035-1036, 'R', 1038-1039, 'K', 1041-1042, 'R', 1044-1045, 'K', 1047-1048, 'R', 1050-1051, 'K', 1053-1054, 'R', 1056-1057, 'K', 1059-1060, 'R', 1062-1063, 'K', 1065-1066, 'R', 1068-1069, 'K', 1071-1072, 'R', 1074-1075, 'K', 1077-1078, 'R', 1080-1081, 'K', 1083-1084, 'R', 1086-1087, 'K', 1089-1090, 'R', 1092-1093, 'K', 1095-1096, 'R', 1098-1099, 'K', 1101-1102, 'R', 1104-1105, 'K', 1107-1108, 'R', 1110-1111, 'K', 1113-1114, 'R', 1116-1117, 'K', 1119-1120, 'R', 1122-1123, 'K', 1125-1126, 'R', 1128-1129, 'K', 1131-1132, 'R', 1134-1135, 'K', 1137-1138, 'R', 1140-1141, 'K', 1143-1144, 'R', 1146-1147, 'K', 1149-1150, 'R', 1152-1153, 'K', 1155-1156, 'R', 1158-1159, 'K', 1161-1162, 'R', 1164-1165, 'K', 1167-1168, 'R', 1170-1171, 'K', 1173-1174, 'R', 1176-117

A:Contents: annotation: X-ray crystallography, 1.9 angstroms  
 C:Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting  
 C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-  
 C:Genetics:  
 A:Gene: GFP  
 A:Introns: 69/3; 167/3  
 C:Superfamily: green-fluorescent protein  
 C:Keywords: chromoprotein; luminescence  
 F:65-67/Cross-link: 5-Imidazolinone (Ser-Gly) #status experimental  
 F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 49.1%; Score 1256; DB 1; Length 238;  
 Best Local Similarity 97.5%; Pred. No. 3.7e-73;  
 Matches 232; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDGYKLTAKFICTTGKLPVWPPTL 60  
 DB 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDGYKLTAKFICTTGKLPVWPPTL 60  
 QY 61 VTTLVGVOCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGFGDGLV 120  
 DB 61 VTTFSGVOCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGFGDGLV 120  
 QY 121 NRIELKIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180  
 DB 121 NRIELKIDFKEDGNILGHKMEYNYNHNYIMADKQNGIKVNFKIRHNIEDGVSQVLAD 180  
 QY 181 HYQONTPIGDGVPVLPDNYHLSQALSQALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238  
 DB 181 HYQONTPIGDGVPVLPDNYHLSQALSQALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238

## RESULT 2

AQJFNV  
 Aequorin precursor - hydromedusa (Aequorea victoria)  
 C:Species: Aequorea victoria  
 C:Date: 28-May-1986 #sequence\_revision 28-May-1986 #text\_change 26-May-2000  
 C:Accession: A03020  
 R:Inouye, S.; Noguchi, M.; Sakaki, Y.; Takagi, Y.; Miyata, T.; Iwanaga, S.; Miyata, T.;  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3154-3158, 1985  
 A:Title: Cloning and sequence analysis of cDNA for the luminescent protein aequorin.  
 A:Reference number: A03020; MUID:85218460; PMID:3858813  
 A:Accession: A03020  
 A:Molecule type: mRNA  
 A:Residues: 1-196 <PRA>  
 A:Cross-references: GB:M11394; NID:g155658; PIDN:AAA27719.1; PID:g155659; GB:L29571; NID:  
 A:Experimental source: Clone A0440  
 C:Comment: The precise function of residues 1-7 is not known.  
 C:Comment: The authors suggest that there are three calcium-binding sites, and that res-  
 ts a blue light.  
 C:Comment: Cysteine residues appear to be critical to the ability of this photoprotein to  
 n, and mercaptoethanol is necessary to regenerate aequorin from apoaequorin.  
 C:Superfamily: calmodulin; calmodulin repeat homology  
 F:8-196/Product: aequorin #status predicted <MA>  
 F:18-50/Domain: calmodulin repeat homology <EF1>  
 F:111-143/Domain: calmodulin repeat homology <EF2>  
 F:147-179/Domain: calmodulin repeat homology <EF3>  
 F:31.33.35.37.42/Binding site: calcium (Asp, Asn, Lys, Glu) #status predicted  
 F:124,126,128,130,135/Binding site: calcium (Asp, Asn, Ala, Glu) #status predicted  
 F:160,162,164,166,171/Binding site: calcium (Asp, Asp, Ser, Gln, Glu) #status predicted

Query Match 40.3%; Score 1032; DB 1; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-59;  
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 SVKLTDFDNPRTWIGRHKHNFELDVNHNKISLIDEMVYKASDIVINNIGATPEQAKRHK 347  
 DB 7 SVKLTDFDNPRTWIGRHKHNFELDVNHNKISLIDEMVYKASDIVINNIGATPEQAKRHK 66  
 QY 348 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRINGDALFDIVDKD 407

DB 67 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRINGDALFDIVDKD 126  
 QY 408 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 467  
 DB 127 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 186  
 QY 468 CEKLYGGAVP 477  
 DB 187 CEKLYGGAVP 196

## RESULT 3

A26623  
 aequorin-1 precursor - hydromedusa (Aequorea victoria)  
 C:Species: Aequorea victoria  
 C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 26-May-2000  
 C:Accession: A26623  
 R:Prasher, D.C.; McCann, R.O.; Longiaru, M.; Cormier, M.J.  
 Biochemistry 26, 1326-1332, 1987  
 A:Title: Sequence comparisons of complementary DNAs encoding aequorin isotypes.  
 A:Reference number: A26623; MUID:87185437; PMID:2882777  
 A:Accession: A26623  
 A:Molecule type: mRNA  
 A:Residues: 1-196 <PRA>  
 A:Cross-references: GB:M16103; NID:g155652; PIDN:AAA27716.1; PID:g155653  
 A:Note: The authors translated the codon GAT for residue 143 as Ala  
 C:Comment: The precise function of residue 1-7 is not known.  
 C:Superfamily: calmodulin; calmodulin repeat homology  
 C:Keywords: calcium binding; EF hand; luminescence  
 F:18-50/Domain: calmodulin repeat homology <EF1>  
 F:111-143/Domain: calmodulin repeat homology <EF2>  
 F:147-179/Domain: calmodulin repeat homology <EF3>

Query Match 37.6%; Score 962; DB 2; Length 196;  
 Best Local Similarity 90.5%; Pred. No. 1.7e-54;  
 Matches 172; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 288 SVKLTDFDNPRTWIGRHKHNFELDVNHNKISLIDEMVYKASDIVINNIGATPEQAKRHK 347  
 DB 7 SVKLTDFDNPRTWIGRHKHNFELDVNHNKISLIDEMVYKASDIVINNIGATPEQAKRHK 66  
 QY 348 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRINGDALFDIVDKD 407  
 DB 67 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRINGDALFDIVDKD 126  
 QY 408 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 467  
 DB 127 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 186  
 QY 468 CEKLYGGAVP 477  
 DB 187 CEKLYGGAVP 196

## RESULT 4

S39022  
 Mitrocomin precursor - Mitrocoma cellularia  
 C:Species: Mitrocoma cellularia  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 26-May-2000  
 C:Accession: S39022  
 R:Fagan, T.F.; Ohmiya, Y.; Blinks, J.R.; Inouye, S.; Tsuji, F.I.  
 FEBS Lett. 333, 301-305, 1993  
 A:Title: Cloning, expression and sequence analysis of cDNA for the Ca(2+)-binding pro-  
 tein mitrocomin from Mitrocoma cellularia.  
 A:Reference number: S39022; MUID:94039781; PMID:8224198  
 A:Accession: S39022  
 A:Molecule type: mRNA  
 A:Residues: 1-198 <PAG>  
 A:Cross-references: EMBL:L31623; NID:g468907; PIDN:AAA29298.1; PID:g468908  
 C:Superfamily: calmodulin; calmodulin repeat homology  
 C:Keywords: calcium binding; EF hand; luminescence  
 F:19-51/Domain: calmodulin repeat homology <EF1>  
 F:112-144/Domain: calmodulin repeat homology <EF2>  
 F:148-180/Domain: calmodulin repeat homology <EF3>



Query Match	29.08;	Score 742;	DB 2;	Length 198;
Best Local Similarity	67.94;	Pred. No. 1.8e-40;		
Matches 129;	Conservative	29;	Mismatches 32;	Indels 0;
Gaps 0;				
QY	288	SVKLTSDFNPRWIGRIKHHMFNFDLVNHNKGISLDEMVIKASDIVINNIGATPEQAKRHK	347	
DB	8	AVKLRTDFDNPKWIAKHHMFNFDLVNHNKGISLDEMVIKASDIVINNIGATPEQAKRHK	347	
QY	348	DAVEAFGGAGMYGVETDPAYIEGWKKLATDELEKYAKNEPILIRIWGDALFDIVDKD	407	
DB	68	KCYEDFGGAGLEYDKDITWPEYIEGWKKLATDELEKSKNQVTLIRIWGDALFDIVDKD	127	
QY	408	ONGAITLDWKATYKKAAGIIQSSDCETETRVCDIDESGOLDVDEMTROHLGFWYTMDDPA	467	
DB	128	RNGSVSLDEWQIYTHCAGIQSQCEATFAKCDLDGDKLDVDEMTROHLGFWYTMDDPA	187	
QY	468	CEKLYGGAVP 477		
DB	188	CEKLYGGAVP 197		
RESULT 5				
S28860				
C:clitin - hydromedusa (Clytia gregarium)				
C:Species: Clytia gregarium				
C:Date: 19-Mar-1997	#sequence_revision 19-Mar-1997	#text_change 21-Jul-2000		
C:Accession: S28860				
R:inouye, S.; Tsujii, F. I.				
FEBS Lett. 315, 343-346, 1993				
A:Title: Cloning and sequence analysis of cDNA for the Ca(2+)-activated photop				
A:Reference number: S28860; MUID:93138101; PMID:8422928				
A:Accession: S28860				
A>Status: preliminary				
A:Molecule type: mRNA				
A:Residues: 1-198 <INO>				
A:Cross-references: GB:U13247; NID:g469233; PIDN:AAA28293.1; PID:g469234				
C:Superfamily: calmodulin; calmodulin repeat homology				
C:Keywords: EF hand; luminescence				
F:20-52/Domain: calmodulin repeat homology <EF1>				
F:113-145/Domain: calmodulin repeat homology <EF2>				
F:149-181/Domain: calmodulin repeat homology <EF3>				
Query Match	25.98;	Score 663;	DB 2;	Length 198;
Best Local Similarity	61.64;	Pred. No. 2e-35;		
Matches 117;	Conservative	36;	Mismatches 37;	Indels 0;
Gaps 0;				
QY	288	SVKLTSDFNPRWIGRIKHHMFNFDLVNHNKGISLDEMVIKASDIVINNIGATPEQAKRHK	347	
DB	9	AVKLRFNFDNPKWIAKHHMFNFDLVNHNKGISLDEMVIKASDIVINNIGATPEQAKRHK	347	
QY	348	DAVEAFGGAGMYGVETDPAYIEGWKKLATDELEKYAKNEPILIRIWGDALFDIVDKD	407	
DB	69	DAVEAFKKIGMDYKGEVEFPFADYGVKELANYDLKLSQKNSLRDWEAGVDFDKD	128	
QY	408	ONGAITLDWKATYKKAAGIIQSSDCETETRVCDIDESGOLDVDEMTROHLGFWYTMDDPA	467	
DB	129	GSGLSLDEWKATGRISGICSSDEDAKTFKCDLDNSGDKLDVDEMTROHLGFWYTMDDPA	188	
QY	468	CEKLYGGAVP 477		
DB	189	ADGLYGNFVP 198		

RESULT 6  
C84470  
hypothetical protein At2g05580 [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 01-Mar-2002  
C:Accession: C84470  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

[illegible]

;Title: Gene organization of the small subunit of human calcium-activated neutral protease  
 ;Accession: A93648; MUID:87066759; PMID:3024120  
 ;Molecule type: DNA  
 ;Residues: 1-268 <MIY>  
 ;Cross-references: GB:M31502  
 ;Ohno, S.; Emori, Y.; Suzuki, K.  
 ;Nucleic Acids Res. 14, 5559, 1986  
 ;Title: Nucleotide sequence of a cDNA coding for the small subunit of human calcium-dep  
 ;Accession: A93631; MUID:86286563; PMID:3016651  
 ;Molecule type: mRNA  
 ;Residues: 1-268 <OHN>  
 ;Cross-references: EMBL:X04106; NID:g35327; PIDN:CAA27726.1; PID:g35328  
 ;Genetics:  
 ;Gene: GDB:CAPN4  
 ;Cross-references: GDB:119752; OMIM:114170

127 GIDFKEDGNTLGHKLFFYNSHNVMADKOKNGIKAMFKIRHNTEDCGSVLADHYQONT 186  
 :  
 749 GLQFEANKVIDQAQOEDEDKENTNLAKLTAVQRKFGTTFE----SVVAEADFSGWT 804  
 :  
 187 PIGDGPVLLPDN---HLYSTQSALSKDPNEKRDMVLLFVTAAGITHGMDELVKSGSG 243  
 :  
 805 NKFNGRTKIDISQGVNAITYQS--SSKKPPPSDFIDIENDPTLGPT-----GSSGGG 855  
 :  
 244 SGQSGGSGSGSQSGSGSGSGSGSGG-----SGSGGSGSGSGSGSG 282



Db 489 RGGSGSYGGGGGGGCGGSGRGGSGGSGGGSGGGSGG 530

## RESULT 15

T31611

hypothetical protein Y50E8A.g - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T31611

R:Steward, C.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z21047

A:Accession: T31611

A&gt;Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1585 &lt;WIL&gt;

A:Cross-references: EMBL:AL117200; NID:e1549770; PIDN:CAB55050.1; CESP:Y50E8A.g

A:Experimental source: clone Y50E8A

C:Genetics:

A:Gene: CESP:Y50E8A.g

A:Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

## Query Match

Best Local Similarity 6.0%; Score 154; DB 2; Length 1585;

Matches 51; Conservative 26; Mismatches 46; Indels 96; Gaps 8;

QY 131 KEDGNLGHKLEYNSHNHYIMADQKNGIKANFKIRHNIEDSGVOLADHYQONTPIGD 190

Db 660 EEDTRQLPHRRHHHHHH-----QLQLQLQHLED-----T 691

QY 191 GPVLLPDNHYLSTOSALSKDPNEKRDHMYLLEFVTAA-----GITHGMDEL----- 236

Db 692 HPEEVEDRH-----EVMQLQEDDAEDDLVVVEEDTRQLPHRRHHHQLQHLEDTHPEST 747

QY 237 -----YKSGSGSGSGGSGGSGG-----SGG 255

Db 748 IWRIRIRGGSGGGYASGGGGGGSGGSGGGRYSSAAAPPPPPPPAPAPAPSSGG 807

QY 256 QSGGS-----GSGGSGSGSGSGGSGGSGG--SGGQSG 285

Db 808 YSGSGSGDSAGCGGSGRGSYSESGSGSGSGSGSGSGSGSG 846

Search completed: June 17, 2003, 16:44:22

Job time : 47.2562 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:31:10 ; Search time 23.5332 Seconds  
(without alignments)  
840.693 Million cell updates/sec

Title: US-09-863-901-5  
Perfect score: 2560  
Sequence: 1 MSKGEELFTGVPIVLVDG.....LGFWTMDPACELKLYGGAVP 477

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1262	49.3	238	1	GFP_AEQVI
2	1032	40.3	196	1	AEQ2_AEQVI
3	962	37.6	196	1	AEQ1_AEQVI
4	742	29.0	198	1	MYTR_MITCE
5	718	28.0	195	1	OBL_OBELO
6	663	25.9	198	1	CLYT_CLYGR
7	171.5	6.7	532	1	ZIC2_HUMAN
8	161.5	6.3	450	1	SWPI_ENCCU
9	161	6.3	115	1	SERL_GALME
10	158.5	6.2	268	1	CANS_HUMAN
11	155	6.1	641	1	EBN1_EBV
12	154.5	6.0	622	1	KICI_HUMAN
13	151.5	5.9	481	1	LORI_MOUSE
14	149.5	5.8	204	1	CORA_MEDSA
15	149	5.8	269	1	CANS_MOUSE
16	149	5.8	421	1	BRJA_MOUSE
17	148.5	5.8	627	1	K2C1_MOUSE
18	148	5.8	1380	1	DDX9_MOUSE
19	148	5.8	1901	1	YZ08_MYCTU
20	147.5	5.8	165	1	GRPI_ORYSA
21	147.5	5.8	434	1	CORA_BPIKE
22	147	5.7	526	1	KICI_BOVIN
23	147	5.7	569	1	KICI_MOUSE
24	146.5	5.7	266	1	CANS_RABIT
25	146	5.7	465	1	GRP2_PHAVU
26	145.5	5.7	166	1	K2C5_BOVIN
27	145	5.7	263	1	CANS_BOVIN
28	144.5	5.6	266	1	CANS_PIG
29	144	5.6	183	1	GRP2_ORYSA
30	143.5	5.6	266	1	CANS_RAT
31	143.5	5.6	434	1	COAA_BPI22
32	143.5	5.6	593	1	KICI_HUMAN
33	143.5	5.6	677	1	SP87_DICDI

## ALIGNMENTS

### RESULT 1

ID	GFP_AEQVI	STANDARD;	PRT;	238 AA.
AC	P42212; Q17104;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cormier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent			
RT	protein.";			
RL	Gene III:229-233(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and			
RT	fluorescence characteristics of the recombinant protein.";			
RL	FEBS Lett. 341:277-280(1994).			
[3]				
RN	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea			
RL	green-fluorescent protein.";			
RL	Biochemistry 32:1212-1218(1993).			
[4]				
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96355665; PubMed=8703075;			
RA	Ormo M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,			
RA	Remington S.J.;			
RT	"Crystal structure of the Aequorea victoria green fluorescent			
RT	protein.";			
RL	Science 273:1392-1395(1996).			
[5]				
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98294543; PubMed=9631087;			
RA	Yang F., Moss L.G., Phillips G.N. Jr.;			
RT	"The molecular structure of green fluorescent protein.";			
RL	Nat. Biotechnol. 14:1246-1251(1996).			
[6]				
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.			
RX	MEDLINE=98455509; PubMed=9782051;			
RA	Wachter R.M., Eisliger M.A., Kallio K., Hanson G.T., Remington S.J.;			
RT	"Structural basis of spectral shifts in the yellow-emission variants			
RT	of green fluorescent protein.";			
RL	Structure 6:1267-1277(1998).			

P29041 xanthomonas  
Q9ng98 drosophila  
P04934 plasmodium  
P50495 plasmodium  
Q00577 homo sapien  
O62233 mus musculu  
P23490 homo sapien  
P09789 petunia hyb  
P03661 bacterioph  
P03662 bacterioph  
Q9jmb3 mus musculu  
P34308 caenorhabdi

34 141.5 5.5 759 1 GSPD\_XANCP  
35 141.5 5.5 1250 1 TP3A\_DROME  
36 141 5.5 1726 1 MSP1\_PLAFC  
37 141 5.5 1726 1 MSP1\_PLAFC  
38 140 5.5 322 1 PUR\_HUMAN  
39 140 5.5 333 1 SIX3\_MOUSE  
40 139.5 5.4 316 1 LORI\_HUMAN  
41 139.5 5.4 384 1 GRP1\_PETHY  
42 139 5.4 424 1 COAA\_BPFD  
43 139 5.4 424 1 COAA\_BPFD  
44 138.5 5.4 148 1 CLM4\_MOUSE  
45 138.5 5.4 734 1 YKR2\_CAEEL

[7]  
 RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RP MEDLINE-99238303; PubMed-10220315;  
 RA Elsiger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;  
 RT "Structural and spectral response of green fluorescent protein  
 variants to changes in pH.";  
 RL Biochemistry 38:5296-5301(1999).  
 CC -!- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSDUCE THE  
 BLUE CHEMILUMINESCENCE OF THE PROTEIN Aequorin INTO GREEN  
 FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO UPON  
 RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PHOTOPROTEIN Aequorin.  
 CC ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER  
 CC ABSORBANCE PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM  
 CC PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.  
 CC -!- SUBUNIT: MONOMER.  
 CC -!- TISSUE SPECIFICITY: PHOTOCYTES.  
 CC -!- PIM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED  
 CC OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON  
 CC CYCLIZATION OF THE RESIDUES SER-DEHYDROTYR-GLY.  
 CC -!- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making  
 CC chimeric proteins of GFP linked to other proteins where it  
 CC functions as a fluorescent protein tag. GFP tolerates N-and C-  
 CC terminal fusion to a broad variety of proteins. It has been  
 CC expressed in bacteria, yeast, slime mold, plants, drosophila,  
 CC zebrafish, and in mammalian cells. As a noninvasive fluorescent  
 CC marker in living cells, it allows for a wide range of applications  
 CC where it may function as a cell lineage tracer, reporter of gene  
 CC expression, or as a measure of protein-protein interactions.  
 CC -!- DATABASE: NAME-Protein Spotlights;  
 CC NOTE-Issue 11 of June 2001;  
 CC WWW="http://www.expasy.org/spotlight/articles/split011.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 DR EMBL; M62654; AAA27722.1; -;  
 DR EMBL; M62653; AAA27721.1; -;  
 DR EMBL; L29345; AAA58246.1; -;  
 DR PDB; 1GFL; 11-JAN-97.  
 DR PDB; 1EMA; 08-NOV-96.  
 DR PDB; 1EMB; 16-JUN-97.  
 DR PDB; 1EMC; 20-AUG-97.  
 DR PDB; 2EMD; 20-AUG-97.  
 DR PDB; 1EME; 20-AUG-97.  
 DR PDB; 1EMF; 20-AUG-97.  
 DR PDB; 1EMG; 12-MAY-99.  
 DR PDB; 1EMK; 20-AUG-97.  
 DR PDB; 1EML; 20-AUG-97.  
 DR PDB; 1EMN; 20-AUG-97.  
 DR PDB; 2EMO; 20-AUG-97.  
 DR PDB; 1BFP; 07-JUL-97.  
 DR PDB; 1YFP; 28-OCT-98.  
 DR InterPro; IPR000786; Green\_fl\_protein.  
 DR Pfam; PF01353; GFP; 1.  
 DR PRINTS; PR01229; GFP; 1.  
 DR PRODOM; PD013756; Green-fl\_protein; 1.  
 KW Luminescence; 3D-structure.  
 FT SITE 65 67 MODIFIED TO FORM THE CHROMOPHORE.  
 FT VARIANT 100 100 F -> Y.  
 FT VARIANT 108 108 T -> S.  
 FT VARIANT 141 141 L -> M.  
 FT VARIANT 219 219 V -> I.  
 FT CONFLICT 25 25 H -> Q (IN REF. 2).  
 FT CONFLICT 157 157 Q -> P (IN REF. 2).  
 FT CONFLICT 172 172 E -> K (IN REF. 2).  
 SQ SEQUENCE 238 AA; 26886 MW; EA5AGF21FBFB6E05 CRC64;

Query Match 49.3%; Score 1262; DB 1; Length 238;  
 Best Local Similarity 98.7%; Pred. No. 3e-69;  
 Matches 235; Conservative 1; Mismatches 0; Gaps 0;  
 QY 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGSDATYKLTFLKFCITCTGKLPVPMPTL 60  
 DB 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGSDATYKLTFLKFCITCTGKLPVPMPTL 60  
 QY 61 VTTLTGYVQCFSPRYPDHMKQDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDPFLV 120  
 DB 61 VTTTSYGVCFSRPYDPMKQDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDPFLV 120  
 QY 121 NRIELKGIDFEDKNILGHKLEYNNSHYIMADKQKNGIKAFKIRHNIEDGSVQLAD 180  
 DB 121 NRIELKGIDFEDKNILGHKLEYNNSHYIMADKQKNGIKAFKIRHNIEDGSVQLAD 180  
 QY 181 HYQONTPTIGDPVLLPDNHYLSTOSALSKDNEKRDHMLVLEFVTAAGITIGMDLYK 238  
 DB 181 HYQONTPTIGDPVLLPDNHYLSTOSALSKDNEKRDHMLVLEFVTAAGITIGMDLYK 238  
 RESULT 2  
 AEQ2\_AEQV1 STANDARD; PRT; 196 AA.  
 ID AEQ2\_AEQV1  
 AC P02592;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Aequorin 2 precursor.  
 OS Aequorea victoria (Jellyfish).  
 CC Eukaryota; Metazoa; Chndaria; Hydrozoa; Hydroida; Leptomedusae;  
 CC Aequoreidae; Aequorea.  
 CC NCBI\_TaxID=6100;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE-85215460; PubMed-3858813;  
 RA Inouye S., Noguchi M., Sakaki Y., Takagi Y., Miyata T., Iwanaga S.,  
 RA Miyata T., Tsuji F.I.;  
 RT "Cloning and sequence analysis of cDNA for the luminescent protein  
 RT aequorin.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 82:3154-3158(1985).  
 CC [2]  
 CC SEQUENCE OF 9-185 FROM N.A. (AEQUORIN 2 AND 3).  
 CC MEDLINE-87185437; PubMed-2882777;  
 RA Prasher D.C., McCann R.O., Longiaru M., Cormier M.J.;  
 RT "Sequence comparisons of complementary DNAs encoding aequorin  
 RT isotypes.";  
 RT Biochemistry 26:1326-1332(1987).  
 CC [3]  
 CC SEQUENCE OF 8-196.  
 CC MEDLINE-86077721; PubMed-2866797;  
 RA Charbonneau H., Walsh K.A., McCann R.O., Prendergast F.G.,  
 RA Cormier M.J., Vanaman T.C.;  
 RT "Amino acid sequence of the calcium-dependent photoprotein aequorin.";  
 RT Biochemistry 24:6762-6771(1985).  
 CC [4]  
 CC MUTAGENESIS.  
 RA Tsuji F.I., Inouye S., Goto T., Sakaki Y.;  
 RT "Site-specific mutagenesis of the calcium-binding photoprotein  
 RT aequorin.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 83:8107-8111(1986).  
 CC [5]  
 CC MUTAGENESIS OF PRO-196.  
 CC MEDLINE-92111761; PubMed-1765170;  
 RA Nomura M., Inouye S., Ohmiya Y., Tsuji F.I.;  
 RT "A C-terminal proline is required for bioluminescence of the Ca(2+)-  
 RT binding photoprotein, aequorin.";  
 RT FEBS Lett. 295:63-66(1991).  
 CC [6]  
 CC DISULFIDE BOND.  
 RA MEDLINE-94009705; PubMed-8405461;  
 RA Ohmiya Y., Kuroono S., Ohashi M., Fagan T.F., Tsuji F.I.;  
 RT "Mass spectrometric evidence for a disulfide bond in aequorin

regeneration.":  
 FEBS Lett. 332:226-228(1993).  
 [7]  
 X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 MEDLINE-20289815; PubMed-10830969;  
 Head J.F., Inouye S., Teranishi K., Shimomura O.;  
 "The crystal structure of the photoprotein aequorin at 2.3-A  
 resolution.":  
 Nature 405:372-376(2000).  
 -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS  
 AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM  
 ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,  
 COELENTERAZINE INTO COELENTERAMIDE AND CO(2) WITH THE  
 CONCOMITANT EMISSION OF LIGHT.  
 -1- PTM: THE REDUCTION OF THE DISULFIDE BOND IS NECESSARY TO  
 REGENERATE AEQUORIN FROM APOAEQUORIN.  
 -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. AEQUORIN FAMILY.  
 -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: L29571; AAA27720.1; -  
 EMBL: M16104; AAA27717.1; -  
 EMBL: M16105; AAA27718.1; -  
 EMBL: M11394; AAA27719.1; -  
 PIR: A03020; A03FNV.  
 PDB: 1EJ3; 3I-MAY-00.  
 InterPro: IPR002048; EF-hand.  
 Pfam: PF00036; efhand; 3.  
 ProDom: PD000012; EF-hand; 1.  
 SMART: SM00054; Eph; 2.  
 PROSITE: PS00018; EF\_HAND; 3.  
 photoprotein; Calcium-binding; Luminescence; Repeat; 3D-structure.  
 KW  
 FT CHAIN 1 7  
 FT SITE 8 196  
 FT SITE 47 57  
 FT SITE 62 72  
 FT SITE 107 117  
 FT CA\_BIND 31 42  
 FT DOMAIN 72 88  
 FT CA\_BIND 124 135  
 FT CA\_BIND 160 171  
 FT DISULFID 152 159  
 FT SITE 196 196  
 FT VARIANT 70 71  
 FT VARIANT 164 164  
 FT MUTAGEN 36 36  
 FT MUTAGEN 129 129  
 FT MUTAGEN 165 165  
 FT MUTAGEN 65 65  
 FT MUTAGEN 152 152  
 FT MUTAGEN 152 152  
 FT MUTAGEN 159 159  
 FT MUTAGEN 187 187  
 FT CONFLICT 37 37  
 FT SEQUENCE 196 AA; 22285 MW; 532DC7A9D29BA80C CRC64;  
 Query Match 40.3%; Score 1032; DB 1; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-55;  
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 288 SVLTSDFNPRIGRKHMFNLDVNHNGKISLDEWYKASDIVINNLGATPEQAKRHK 347  
 Db 7 SVLTSDFNPRIGRKHMFNLDVNHNGKISLDEWYKASDIVINNLGATPEQAKRHK 66  
 QY 348 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLRINGWGDALFDVVDK 407  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 67 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLRINGWGDALFDVVDK 126  
 QY 408 QNGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFYWTMDPA 467  
 Db 127 QNGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFYWTMDPA 186  
 QY 468 CEKLYGGAVP 477  
 Db 187 CEKLYGGAVP 196  
 |||||||||||  
 RESULT 3  
 AEQI\_AEQVI  
 ID AEQI\_AEQVI STANDARD; PRT; 196 AA.  
 AC P07164;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Aequorin 1 precursor.  
 OS Aequorea victoria (Jellyfish).  
 OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;  
 OC Aequoreidae; Aequorea.  
 OX NCBI\_TaxID=6100;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-87185437; PubMed-2882777;  
 RA Prasher D.C., McCann R.O., Longiaru M., Cormier M.J.;  
 RT "Sequence comparisons of complementary DNAs encoding aequorin  
 RT isotypes.":  
 RL Biochemistry 26:1326-1332(1987).  
 RN [2]  
 RP SEQUENCE OF 8-196.  
 RX MEDLINE-8607721; PubMed-2866797;  
 RA Charbonneau H., Walsh K.A., McCann R.O., Prendergast F.G.,  
 RA Cormier M.J., Vanaman T.C.;  
 RT "Amino acid sequence of the calcium-dependent photoprotein aequorin.":  
 RL Biochemistry 24:6762-6771(1985).  
 RN [3]  
 RP MUTAGENESIS OF PRO-196.  
 RX MEDLINE-92111761; PubMed-1765170;  
 RA Nomura M., Inouye S., Ohmiya Y., Tsuji F.I.;  
 RT "A C-terminal proline is required for bioluminescence of the Ca(2+)-  
 RT binding photoprotein, aequorin.":  
 RL FEBS Lett. 295:63-66(1991).  
 RN [4]  
 RP DISULFIDE BOND.  
 RX MEDLINE-94009705; PubMed-8405461;  
 RA Ohmiya Y., Kurono S., Ohashi M., Fagan T.F., Tsuji F.I.;  
 RT "Mass spectrometric evidence for a disulfide bond in aequorin  
 RT regeneration.":  
 RL FEBS Lett. 332:226-228(1993).  
 CC -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS  
 AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM  
 ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,  
 COELENTERAZINE INTO COELENTERAMIDE AND CO(2) WITH THE  
 CONCOMITANT EMISSION OF LIGHT.  
 CC -1- PTM: THE REDUCTION OF THE DISULFIDE BOND IS NECESSARY TO  
 REGENERATE AEQUORIN FROM APOAEQUORIN.  
 CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. AEQUORIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: M16103; AAA27716.1; -  
 PIR: A26623; A26623.  
 HSP: P02592; 1EJ3.  
 InterPro: IPR002048; EF-hand.  
 DR

DR Pfam; PF00036; ehand; 3.  
 DR PRODOM; PD000012; EF-hand; 1.  
 DR SMART; SM00054; EFh; 2.  
 DR PROSITE; PS00018; EF-HAND; 3.  
 KW Photoprotein; Calcium-binding; Luminescence; Repeat.  
 FT PROPEP 1 7  
 FT CHAIN 8 196  
 FT SITE 47 57  
 FT SITE 62 72  
 FT SITE 107 117  
 FT CA-BIND 31 42  
 FT DOMAIN 72 88  
 FT CA-BIND 124 135  
 FT CA-BIND 160 171  
 FT DISULFID 152 159  
 SQ SEQUENCE 196 AA; 22514 MW; 9AA5B636288A5B8F CRC64;  
 Query Match 37.6%; Score 962; DB 1; Length 196;  
 Best Local Similarity 90.5%; Pred. No. 2.3e-51;  
 Matches 172; Conservative 14; Mismatches 4; Indels 0; Gaps 0;  
 QY 288 SVKLTDFDNPWGRHKHMFNFVDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347  
 Db 7 SVKLTDFDNPWGRHKHMFNFVDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 66  
 QY 348 DAVEAFGGAGMKYGVETDPAYIEGWKKLATDELEKAKNEPTLIRIWGDALFDIVDKD 407  
 Db 67 DAVEAFGGAGMKYGVETDPAYIEGWKKLATDELEKAKNEPTLIRIWGDALFDIVDKD 126  
 QY 408 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESSGOLDVDVDMTROHGLGFWYTMDDPA 467  
 Db 127 QNGAISLDEWKATYKSDGIIQSSDCEETFRVCDIDESSGOLDVDVDMTROHGLGFWYTMDDPA 186  
 QY 468 CEKLYGGAVP 477  
 Db 187 CEKLYGGAVP 196  
 RESULT 4  
 MYTR\_MITCE STANDARD; PRT; 198 AA.  
 AC P39047;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mitrocomin precursor.  
 GN M17.  
 OS Mitrocoma cellullaria (Halistaura mitrocoma).  
 OC Eukaryota; Metazoa; Chndaria; Hydrozoa; Hydroids; Anthomedusae;  
 OC Mitrocomidae; Mitrocoma.  
 OX NCBI\_TaxID=31874;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94039781; PubMed=82241198;  
 RA Fagan T.F., Ohmiya Y., Blinks J.R., Inouye S., Tsuji F.I.;  
 RT Cloning, expression and sequence analysis of cDNA for the Ca(2+)-  
 binding photoprotein, mitrocomin.";  
 RL FEBS Lett. 333:301-305(1993).  
 CC -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS  
 AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM  
 ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,  
 COLENTHERAZINE INTO COLENTHERAMIDE AND CO(2) WITH THE  
 CONCOMITANT EMISSION OF LIGHT.  
 CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. Aequorin FAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L31623; AAA29298.1; -  
 DR PIR; S39022; S39022.  
 DR HSP; P02592; IEJ3.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00036; ehand; 3.  
 DR PRODOM; PD000012; EF-hand; 1.  
 DR SMART; SM00054; EFh; 2.  
 DR PROSITE; PS00018; EF-HAND; 3.  
 KW Photoprotein; Calcium-binding; Luminescence; Repeat.  
 FT PROPEP 1 8  
 FT CHAIN 9 198  
 FT CA-BIND 32 43  
 FT DOMAIN 73 89  
 FT CA-BIND 125 136  
 FT CA-BIND 161 172  
 FT DISULFID 153 160  
 SQ SEQUENCE 198 AA; 22714 MW; 8F6307EF0966F670 CRC64;  
 Query Match 29.0%; Score 742; DB 1; Length 198;  
 Best Local Similarity 67.9%; Pred. No. 3.6e-38;  
 Matches 129; Conservative 29; Mismatches 32; Indels 0; Gaps 0;  
 QY 288 SVKLTDFDNPWGRHKHMFNFVDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347  
 Db 8 AVKLTDFDNPWGRHKHMFNFVDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 67  
 QY 348 DAVEAFGGAGMKYGVETDPAYIEGWKKLATDELEKAKNEPTLIRIWGDALFDIVDKD 407  
 Db 68 KCVDFGGAGLEVDKDTTPPEYIEGWKKLATDELEKAKNEPTLIRIWGDALFDIVDKD 127  
 QY 408 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESSGOLDVDVDMTROHGLGFWYTMDDPA 467.  
 Db 128 RGVSLDEWITQTHCAGIQSSRQCCEATFAHCDLGDGKLDVDMTROHGLGFWYTMDDPA 187  
 QY 468 CEKLYGGAVP 477  
 Db 188 CEKLYGGAVP 197  
 RESULT 5  
 OBL\_OBELO STANDARD; PRT; 195 AA.  
 AC Q27709;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Obelin precursor (OBL).  
 OS Obelia longissima (Black sea hydrosoma).  
 OC Eukaryota; Metazoa; Chndaria; Hydrozoa; Hydroids; Leptomedusae;  
 OC Campanulariidae; Obelia.  
 OX NCBI\_TaxID=32570;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95180731; PubMed=7875600;  
 RA Illarionov B.A., Bondar V.S., Illarionova V.A., Vysotski E.S.;  
 RT "Sequence of the cDNA encoding the Ca(2+)-activated photoprotein  
 Obelin from the hydroid polyp Obelia longissima.";  
 RL Gene 153:273-274(1995).  
 CC -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS  
 AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM  
 ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,  
 COLENTHERAZINE INTO COLENTHERAMIDE AND CO(2) WITH THE  
 CONCOMITANT EMISSION OF LIGHT.  
 CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. Aequorin FAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```

CC EMBL: U07128; AAA67708.1; -.
CC HSSP: P02592; 1EJ3.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ehand; 3.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF-HAND; 3.
KW Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 6 POTENTIAL.
FT CHAIN 7 195 OBELIN.
FT CA_BIND 30 41 EF-HAND 1 (POTENTIAL).
FT DOMAIN 71 87 ANCESTRAL CALCIUM SITE 2.
FT CA_BIND 123 134 EF-HAND 3 (POTENTIAL).
FT CA_BIND 159 170 EF-HAND 4 (POTENTIAL).
FT DISULFID 151 158 BY SIMILARITY.
SQ SEQUENCE 195 AA; 22226 MW; 5D002270B73D3663 CRC64;

Query Match 28.0%; Score 718; DB 1; Length 195;
Best Local Similarity 68.4%; Pred. No. 9.8e-37;
Matches 130; Conservative 23; Mismatches 37; Indels 0; Gaps 0;

QY 288 SVKLTSDFDNPRWIGRHKHMFNFDVNHNGKISLDEWYKASDIVINNIGLGPATPEQAKRHK 347
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 6 AVKLKTFDNPWKIKRHKHMFNFDLNGKTKITLDEIVSKASDDICAKLEATPEQTKRHQ 65
QY 348 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFIDVKD 407
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 66 VCVEAFPGCGMEYGEIAFPQFLDQWKLATSELKWNARNEPTLIRENGDAVFDFDKD 125
QY 408 QNGAITLDEWKATKAAGIIQSSDCEETFRVCDIDESQGLDYDENTROHLGFWYTMDDPA 467
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 126 GSGTITLDEWKAYKRGISGSPSQEDCEATFRHCDLSDGLDYDENTROHLGFWYTLDP 185
QY 468 CEKLYGGAVP 477
DB :|||:|||||:
DB 186 ADGLYGNVGP 195

```

#### RESULT 6

```

ID CLYT_CLYGR STANDARD; PRT; 198 AA.
AC Q08121;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytlin precursor (Phalidin).
OS Cytlin gregaria (Phalididum gregarium).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Campanulariidae; Clytia.
OX NCBI_TaxID=27801;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93138101; PubMed=8422928;
RA Inouye S., Tsuji F.I.;
RT "Cloning and sequence analysis of cDNA for the Ca(2+)-activated
RT photoprotein, clytin.";
RL FEBS Lett. 315:343-346(1993).
CC -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
CC AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
CC ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
CC COELENTERAZINE INFO COELENTERAMIDE AND CO(2) WITH THE
CC CONCOMITANT EMISSION OF LIGHT.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. AEQUORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC

```

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```

CC EMBL: L13247; AAA28293.1; -.
CC DR EMBL: X70221; CAA49754.1; -.
DR PIR: S28860; S28860.
DR HSSP: P02592; 1EJ3.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ehand; 3.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF-HAND; 3.
KW Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 9 POTENTIAL.
FT CHAIN 10 198 CLYTIN.
FT CA_BIND 33 44 EF-HAND 1 (POTENTIAL).
FT DOMAIN 74 90 ANCESTRAL CALCIUM SITE 2.
FT CA_BIND 126 137 EF-HAND 3 (POTENTIAL).
FT CA_BIND 162 173 EF-HAND 4 (POTENTIAL).
FT DISULFID 148 161 BY SIMILARITY.
SQ SEQUENCE 198 AA; 23541 MW; 23F1E399667F9059 CRC64;

Query Match 25.9%; Score 663; DB 1; Length 198;
Best Local Similarity 61.6%; Pred. No. 2e-33;
Matches 117; Conservative 36; Mismatches 37; Indels 0; Gaps 0;

QY 288 SVKLTSDFDNPRWIGRHKHMFNFDVNHNGKISLDEWYKASDIVINNIGLGPATPEQAKRHK 347
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 9 AVKLKTFDNPWKIKRHKHMFNFDLNGKTKITLDEIVSKASDDICAKLEATPEQTKRHQ 68
QY 348 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFIDVKD 407
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 69 DAVEAFKGIQMDYIGREVEFPFADGWKELAYDLKLSQNKSLINDWGEAVDFIDFKD 128
QY 408 QNGAITLDEWKATKAAGIIQSSDCEETFRVCDIDESQGLDYDENTROHLGFWYTMDDPA 467
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 129 GSGSILDEWKAYGRISGICSDDEAKTFKHCDLSDGLDYDENTROHLGFWYTLDPN 188
QY 468 CEKLYGGAVP 477
DB :|||:|||||:
DB 189 ADGLYGNVGP 198

```

#### RESULT 7

```

ID ZIC2_HUMAN STANDARD; PRT; 532 AA.
AC Q95409; Q9H309;
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein ZIC2 (zinc finger protein of the cerebellum 2).
GN ZIC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT HPES POLY-ALA INSERTION.
RX MEDLINE=98442655; PubMed=9771712;
RA Brown S.A., Warburton D., Brown L.Y., Yu C.Y., Roeder E.R.,
RA Stengel-Rutkowski S., Hennekam R.C., Muenke M.;
RT "Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila
RT odd-paired.";
RL Nat. Genet. 20:180-183(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20556339; PubMed=10984499;
RA Yang Y., Hwang C.K., Junn E., Lee G., Mouradian M.M.;
RT "ZIC2 and Sp3 repress Sp1-induced activation of the human D1A dopamine
RT receptor gene.";
RL J. Biol. Chem. 275:38863-38869(2000).
CC -1- SURCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: DEFECTS IN ZIC2 ARE A CAUSE OF HOLOPROSENCEPHALY TYPE 5

```









CC entitles requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

DR EMBL; M34398; AAA39444.1; -  
DR EMBL; U09189; AAA82152.1; -  
DR PIR; A35628; A35628.  
DR HSP; P02876; 9WCA.  
DR MGD; MGI:96816; Lor.  
KW Keratinization.  
SQ SEQUENCE 481 AA; 37830 MW; 97349A786FF239FE CRC64;

Query Match 5.9%; Score 151.5; DB 1; Length 481;  
Best Local Similarity 45.2%; Pred. No. 0.024; 33; Indels 11; Gaps 3;  
Matches 38; Conservative 2; Mismatches 3;  
QY 228 GITHGMDLYKSGSG-SGGQSGSGG-----SGGQSGSGG-----SGGQSGSGGSGGSGG 276  
DB 395 GSSGCGGXYSGGCGGCGGSGGSGGCGGCGGSGGSGGCGGCGGSGGCGGSGGSGGSGGSGG 454  
QY 277 GSGGCGGSLRSVKLTSDNDPRW 300  
DB 455 GGGSGGGRGVPVCHQTOOQAPTW 478

RESULT 14  
CORA\_MSDA STANDARD; PRT; 204 AA.  
AC Q07202;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE Cold and drought-regulated protein CORA.  
GN CORA.  
OS Medicago sativa (Alfalfa).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
OX NCBI\_TaxID=3879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Apica;  
RX MEDLINE=94143496; PubMed=8310076;  
RA Laberge S., Castonguay V., Vezina L.-P.;  
RT "New cold- and drought-regulated gene from Medicago sativa.";  
RL Plant Physiol. 101:1411-1412(1993)  
CC -!- FUNCTION: MAY BE INVOLVED IN RESISTANCE OF THE PLANT TO  
ENVIRONMENTAL STRESS.  
CC -!- INDUCTION: BY COLD, ABSCISIC ACID (ABA) AND DROUGHT STRESS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entitles requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

DR EMBL; L03708; AAA9833.1; -  
KW Multigene family; Repeat.  
FT DOMAIN 54 176 7 X 6 AA REPEATS OF Y-N-H-G-G-G.  
FT REPEAT 54 59 1-1.  
FT REPEAT 65 70 1-2.  
FT REPEAT 71 76 1-3.  
FT REPEAT 78 83 1-4.  
FT REPEAT 85 90 1-5.  
FT REPEAT 164 169 1-6.  
FT REPEAT 171 176 1-7.  
FT DOMAIN 98 192 11 X 3 AA REPEATS OF H-G-G.  
FT REPEAT 98 100 2-1.  
FT REPEAT 101 103 2-2.  
FT REPEAT 112 114 2-3.  
FT REPEAT 115 117 2-4.  
FT REPEAT 126 128 2-5.

FT REPEAT 129 131 2-6.  
FT REPEAT 178 180 2-7.  
FT REPEAT 181 183 2-8.  
FT REPEAT 184 186 2-9.  
FT REPEAT 187 189 2-10.  
FT REPEAT 190 192 2-11.  
SQ SEQUENCE 204 AA; 19599 MW; 8A2C082359FCC17F CRC64;

Query Match 5.8%; Score 149.5; DB 1; Length 204;  
Best Local Similarity 35.4%; Pred. No. 0.012;  
Matches 35; Conservative 11; Mismatches 44; Indels 9; Gaps 3;

QY 228 GITHGMDLYKSGSG-SGGQSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 283  
DB 70 GYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGG 129  
QY 284 SGLRSVKL-----TSDFDNPRIWIRHMKHMFNFDVNHNG 317  
DB 130 GGAESVAVQTEKTNEVDNDAKYGGGYNNDGRGGYNHGG 168

RESULT 15  
CANS\_MOUSE STANDARD; PRT; 269 AA.  
ID CANS\_MOUSE STANDARD; PRT; 269 AA.  
AC O88456; OSVEKA;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)  
DE (Calcium-activated neutral proteinase) (CANP).  
GN CAPN1 OR CAPN4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20285452; PubMed=10825211;  
RA Arthur J.S.C., Elce J.S., Hegadorn C., Williams K., Greer P.A.;  
RT "Disruption of the murine calpain small subunit gene, Capn4: calpain  
RT is essential for embryonic development but not for cell growth and  
RT division.";  
RL Mol. Cell. Biol. 20:4474-4481(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which  
CC catalyze limited proteolysis of substrates involved in  
CC cytoskeletal remodeling and signal transduction.  
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small  
CC (regulatory) subunit.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma  
CC membrane upon Ca++ binding (By similarity).  
CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entitles requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

DR EMBL; AF058298; AAC97194.1; -  
DR EMBL; BC018352; AAH18352.1; -  
DR HSP; P04632; IKFU.  
DR MGD; MGI:88266; Capnsl.  
DR InterPro; IPR002048; EF-hand.  
DR Pfam; PF00036; efhand; 3.  
DR ProDom; PD000012; EF-hand; 1.  
DR PROSITE; PS00018; EF\_HAND; 2.  
KW Calcium-binding; Repeat.

Search completed: June 17, 2003, 16:41:32  
Job time : 25.5332 secs





Result No.	Score	Query #		Length	DB	ID	Description
		Match	Length				
1	1256	49.1	238	5	Q93125		aequorea vi
2	1254	49.0	238	5	Q27903		Q927903 unidentifie
3	1231	48.1	758	4	Q96701		Q96701 homo sapien
4	1220	47.7	238	5	Q17105		Q17105 aequorea vi
5	1205	47.1	238	5	Q17106		Q17106 aequorea vi
6	1100	43.0	238	5	Q8WTC6		Q8WTC6 aequorea ma
7	1096	42.8	238	5	Q8WTP95		Q8WTP95 aequorea ma
8	1093	42.7	238	5	Q8WTC4		Q8WTC4 aequorea ma
9	1091	42.6	238	5	Q8WTD0		Q8WTD0 aequorea ma
10	1090	42.6	238	5	Q8WTC9		Q8WTC9 aequorea ma
11	1090	42.6	238	5	Q8WTC8		Q8WTC8 aequorea ma
12	1087	42.5	238	5	Q8WTC7		Q8WTC7 aequorea ma
13	1086	42.4	238	5	Q8WTC5		Q8WTC5 aequorea ma
14	915	35.7	195	5	Q8WQY8		Q8WQY8 aequorea pa
15	899	35.1	195	5	Q8WQY7		Q8WQY7 aequorea ma
16	705	27.5	195	5	Q8T620		Q8T620 obelia geni

QV 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTILKFICTTGKLPVPWPPTL 60

```
|||||
Db 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKGLTLKFICTTGKLPVWPPTL 60
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120
Db 61 VTTFGVGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 2
Q27903 PRELIMINARY; PRT; 238 AA.
AC Q27903;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI_TaxID=32644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97299832; PubMed=9154981;
RA Rowland G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
RL Plant Mol. Biol. 33:989-999(1997)."
DR EMBL; X96418; CAAG5278.1; -
DR HSSP; P42212; 1GFL.
DR InterPro; IPR000786; Green_fl_protein.
DR PRINTS; PR01229; GFP.
DR PRODOM; PD013756; Green_fl_protein; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
SQ SEQUENCE 238 AA; 26884 MW; CA932D47262AF2D3 CRC64;

Query Match 49.0%; Score 1254; DB 5; Length 238;
Best Local Similarity 97.9%; Pred. NO. 8.3e-82;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKGLTLKFICTTGKLPVWPPTL 60
Db 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKGLTLKFICTTGKLPVWPPTL 60
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120
Db 61 VTTFGVGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 3
Q96JQ1 PRELIMINARY; PRT; 758 AA.
AC Q96JQ1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ralchu404X.
GN RAICR0404X.

|||||
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21322811; PubMed=11429608;
RA Mochizuki N., Yamashita S., Kurokawa K., Ohba Y., Nagai T.,
RA Miyawaki A., Matsuda M.;
RT "Spacio-temporal Images of Growth Factor-induced Activation of Ras and
RT Rapi. ";
RL Nature 411:1065-1068(2001).
DR EMBL; AB051846; BAB61868.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_trnsfrmg.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01353; GFP; 2.
DR Pfam; PF00071; Ras; 1.
DR Pfam; PF02196; RBD; 1.
DR PRODOM; PD013756; Green_fl_protein; 2.
DR TIGRfams; TIGR00231; small_GTP; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW GTP-binding.
SQ SEQUENCE 758 AA; 85015 MW; 8612408F607CFD49 CRC64;
```

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21322811; PubMed=11429608;
RA Mochizuki N., Yamashita S., Kurokawa K., Ohba Y., Nagai T.,
RA Miyawaki A., Matsuda M.;
RT "Spacio-temporal Images of Growth Factor-induced Activation of Ras and
RT Rapi. ";
RL Nature 411:1065-1068(2001).
DR EMBL; AB051846; BAB61868.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_trnsfrmg.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01353; GFP; 2.
DR Pfam; PF00071; Ras; 1.
DR Pfam; PF02196; RBD; 1.
DR PRODOM; PD013756; Green_fl_protein; 2.
DR TIGRfams; TIGR00231; small_GTP; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW GTP-binding.
SQ SEQUENCE 758 AA; 85015 MW; 8612408F607CFD49 CRC64;

Query Match 48.1%; Score 1231; DB 4; Length 758;
Best Local Similarity 94.3%; Pred. NO. 1.9e-79;
Matches 233; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKGLTLKFICTTGKLPVWPPTL 60
Db 499 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKGLTLKFICTTGKLPVWPPTL 558
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 120
Db 559 VTTLTGWQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLV 618
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
Db 619 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVQLAD 678
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
Db 679 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 738
QY 241 GSGSGGQ 247
Db 739 KMSKDGK 745

RESULT 4
Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -
DR HSSP; P42212; 1GFL.
DR InterPro; IPR000786; Green_fl_protein.
```

DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
FT NON\_TER 238  
SQ SEQUENCE 238 AA; 26E2BE450E748E44 CRC64;  
  
Query Match 47.7%; Score 1220; DB 5; Length 238;  
Best Local Similarity 94.5%; Pred. No. 2.2e-79;  
Matches 225; Conservative 6; Mismatches 7; Indels 0; Gaps 0;  
  
QY 1 MSKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKGLTKFICTTGKLPVPWPTL 60  
DB 1 MSKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKGLTKFICTTGKLPVPWPTL 60  
  
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVFEGETLV 120  
DB 61 VTTSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVFEGETLV 120  
  
QY 121 NRIELKGIDPKEDGNILGHKLEYNSHNHYINADKQKNGIKANFKIRHNIEDGSVOLAD 180  
DB 121 NRIELKGIDPKEDGNILGHKLEYNSHNHYINADKQKNGIKANFKIRHNIEDGSVOLAD 180  
  
QY 181 HYQONTPIGDPVLLPDNHYLSQSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238  
DB 181 HYQONTPIGDPVLLPDNHYLSQSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238  
  
RESULT 5  
Q17106 PRELIMINARY; PRT; 238 AA.  
ID Q17106  
AC Q17106  
DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Green fluorescent protein (Fragment).  
GN GFP.  
OS Asquorea victoria (Jellyfish).  
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=6100;  
RN [1]  
RP STRAIN=J.N., Campbell A.K.;  
RA Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
RL EMBL; X83960; CAA58790.1; -.  
DR HSP; P42212; 1BFP.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR PRINTS; PR01229; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
FT NON\_TER 238  
SQ SEQUENCE 238 AA; 26867 MW; BD4648262D8EABD4 CRC64;  
  
Query Match 47.1%; Score 1205; DB 5; Length 238;  
Best Local Similarity 93.7%; Pred. No. 2.6e-78;  
Matches 223; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
  
QY 1 MSKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKGLTKFICTTGKLPVPWPTL 60  
DB 1 MSKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKGLTKFICTTGKLPVPWPTL 60  
  
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVFEGETLV 120  
DB 61 VTTSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVFEGETLV 120  
  
QY 121 NRIELKGIDPKEDGNILGHKLEYNSHNHYINADKQKNGIKANFKIRHNIEDGSVOLAD 180  
DB 121 NRIELKGIDPKEDGNILGHKLEYNSHNHYINADKQKNGIKANFKIRHNIEDGSVOLAD 180  
  
QY 181 HYQONTPIGDPVLLPDNHYLSQSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238  
DB 181 HYQONTPIGDPVLLPDNHYLSQSALSQSDPHKRDHMLLEFVTSAGITHGMDELYK 238

RESULT 6  
Q8WTC6 PRELIMINARY; PRT; 238 AA.  
ID Q8WTC6  
AC Q8WTC6  
DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Green fluorescent protein.  
GN GFP.  
OS Asquorea macrodactyla.  
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP STRAIN=GFPX10V;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,  
Li S.J., Xia N.S.;  
RT \*Colorful mutants of green fluorescent protein from Aequorea  
macrodactyla.\*;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF435431; AAL33916.1; -.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.  
SQ SEQUENCE 238 AA; 688FD75E88926903 CRC64;  
  
Query Match 43.0%; Score 1100; DB 5; Length 238;  
Best Local Similarity 82.8%; Pred. No. 8.1e-71;  
Matches 197; Conservative 20; Mismatches 21; Indels 0; Gaps 0;  
  
QY 1 MSKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKGLTKFICTTGKLPVPWPTL 60  
DB 1 MSKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKGLTKFICTTGKLPVPWPTL 60  
  
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVFEGETLV 120  
DB 61 VTTSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVFEGETLV 120  
  
QY 121 NRIELKGIDPKEDGNILGHKLEYNSHNHYINADKQKNGIKANFKIRHNIEDGSVOLAD 180  
DB 121 NRIELKGIDPKEDGNILGHKLEYNSHNHYINADKQKNGIKANFKIRHNIEDGSVOLAD 180  
  
QY 181 HYQONTPIGDPVLLPDNHYLSQSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238  
DB 181 HYQONTPIGDPVLLPDNHYLSQSALSQSDPNKRDHMLLEFVTSAGITHGMDELYK 238  
  
RESULT 7  
Q8WP95 PRELIMINARY; PRT; 238 AA.  
ID Q8WP95  
AC Q8WP95  
DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Green fluorescent protein.  
GN GFPX.  
OS Asquorea macrodactyla.  
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP STRAIN=GFPX, AND GFPNAX;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.Q.,  
Li S.J., Xia N.S.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY013824; AAK02062.1; -.  
DR EMBL; AY013821; AAK02059.1; -.  
DR InterPro; IPR000786; Green\_fl\_protein.  
DR Pfam; PF01353; GFP; 1.  
DR ProDom; PD013756; Green\_fl\_protein; 1.

```

SQ SEQUENCE 238 AA; 27049 MW; 8185D0E5E29012B CRC64;
Query Match 42.8%; Score 1096; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.6e-70;
Matches 196; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLCFCTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLCFCTTGKLPVWPPTL 60
QY 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNKTRAEVFEGLTV 120
DB 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNKTRAEVFEGLTV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIKANFKIRHNIEDGSVOLAD 180
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQNTPIGDGPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238
DB 181 HYQNTPIGDGPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 8
Q8WTC4 PRELIMINARY; PRT; 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPX161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL33912.1;
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match 42.7%; Score 1093; DB 5; Length 238;
Best Local Similarity 83.6%; Pred. No. 2.6e-70;
Matches 199; Conservative 16; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLCFCTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLCFCTTGKLPVWPPTL 60
QY 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNKTRAEVFEGLTV 120
DB 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNKTRAEVFEGLTV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQNTPIGDGPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238
DB 181 HYQNTPIGDGPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 9
Q8WTC9 PRELIMINARY; PRT; 238 AA.
AC Q8WTC9;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPX162;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL33913.1;
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

```

```

ID Q8WTC9 PRELIMINARY; PRT; 238 AA.
AC Q8WTC9;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPX161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL33912.1;
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;

Query Match 42.6%; Score 1091; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 3.6e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLCFCTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLVDGDNVGHKFSVSGEGDATYKLTLCFCTTGKLPVWPPTL 60
QY 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNKTRAEVFEGLTV 120
DB 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNKTRAEVFEGLTV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQNTPIGDGPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238
DB 181 HYQNTPIGDGPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 10
Q8WTC9 PRELIMINARY; PRT; 238 AA.
AC Q8WTC9;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPX162;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL33913.1;
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

```

```
Query Match      42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPIVVELDGDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60
DB 1 MSKGEELFTGVPIVVELDGDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60

QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 121 NRIELKGIDFEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180

QY 121 NRIELKGMDPKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGMDPKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180

QY 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

QY 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 11
Q8WTC8      PRELIMINARY;      PRT;      238 AA.
ID AC Q8WTC8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPX1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match      42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPIVVELDGDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60
DB 1 MSKGEELFTGVPIVVELDGDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60

QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 121 NRIELKGIDFEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180

QY 121 NRIELKGMDPKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGMDPKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180

QY 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

QY 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 12
Q8WTC7      PRELIMINARY;      PRT;      238 AA.
ID AC Q8WTC7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPX1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match      42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPIVVELDGDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60
DB 1 MSKGEELFTGVPIVVELDGDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60

QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYQVQERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 121 NRIELKGIDFEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180

QY 121 NRIELKGMDPKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGMDPKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180

QY 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

QY 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238

RESULT 13
Q8WTC5      PRELIMINARY;      PRT;      238 AA.
ID AC Q8WTC5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Orange fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPX1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435432; AAL33917.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27018 MW; 75521EA5534E573A CRC64;

Query Match      42.4%; Score 1086; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 8.1e-70;
```

Matches 198; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

QY 1 MSKGBELTGVVPIILVELDGVNKHKSVSSEGGDATYKLTFLKFKCTTGKLPVWPPTL 60  
DB 1 MSKGBELTGVVPIILVELDGVNKHKSVSSEGGDATYKLTFLKFKCTTGKLPVWPPTL 60  
QY 61 VTTLYGVOQTSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVFEGDTLV 120  
DB 61 VTTLYGVOQTSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVFEGDTLV 120  
QY 121 NRLEKLGIDFKEDGNILGHKLEYNNSHVYIMADKOKNGIKANFKIRHNIEDSGVOLAD 180  
DB 121 NRLEKLGIDFKEDGNILGHKLEYNNSHVYIMADKOKNGIKANFKIRHNIEDSGVOLAD 180  
QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDNEKRDHMLVLEFVTAAGITHGMDELYK 238  
DB 181 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDNEKRDHMLVLEFVTAAGITHGMDELYK 238  
181 HTQVNPVLDGVPVLPINHYLSYQTAISKDRNEYRDHVMVLEFFSACGHTGMDELYK 238

## RESULT 14

Q8WQY8 PRELIMINARY; PRT; 195 AA.  
AC Q8WQY8;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Apoeaquorin.  
GN AEQUORIN.  
OS Aequorea parva.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=148610;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AEQXM;  
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.Q.,  
Li S.J., Xia N.S.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY013822; AAK02060.1; -  
DR InterPro; IPR002048; EF-hand.  
DR Pfam; PF00036; EF-hand; 3.  
DR ProDom; PD000012; EF-hand; 1.  
DR SMART; SM00054; EFh; 3.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_3.  
SQ SEQUENCE 195 AA; 22248 MW; 73EE72E7E0C140BE CRC64;

Query Match 35.7%; Score 915; DB 5; Length 195;  
Best Local Similarity 85.3%; Pred. No. 9.7e-58;  
Matches 162; Conservative 18; Mismatches 10; Indels 0; Gaps 0;

QY 288 SVKLTSDFNPRWIGRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347  
DB 6 AVKLEPDFENPRWIGRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 65  
QY 348 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407  
DB 66 EAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 125  
QY 408 QNGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHGLFWYTMCPA 467  
DB 126 QNGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHGLFWYTMCPA 185  
QY 468 CEKLYGGAVP 477  
DB 186 CEKLYGGAVP 195

## RESULT 15

Q8WQY7 PRELIMINARY; PRT; 195 AA.  
AC Q8WQY7;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Apoeaquorin.  
GN AEQUORIN.  
OS Aequorea macrodactyla.  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;  
OC Aequoreidae; Aequorea.  
OX NCBI\_TaxID=147615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AEQXM;  
RA Luo W.X., Zhang J., Yang H.J., Li S.W., Xie X.Y., Qin Y.X., Pang S.Q.,  
Li S.J., Xia N.S.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY013823; AAK02061.1; -  
DR InterPro; IPR002048; EF-hand.  
DR Pfam; PF00036; EF-hand; 3.  
DR ProDom; PD000012; EF-hand; 1.  
DR SMART; SM00054; EFh; 2.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_3.  
SQ SEQUENCE 195 AA; 22266 MW; A459391B4B8003BF CRC64;

Query Match 35.1%; Score 899; DB 5; Length 195;  
Best Local Similarity 83.7%; Pred. No. 1.4e-56;  
Matches 159; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

QY 288 SVKLTSDFNPRWIGRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347  
DB 6 AVKLEPDFENPRWIGRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 65  
QY 348 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407  
DB 66 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 125  
QY 408 QNGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHGLFWYTMCPA 467  
DB 126 QNGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHGLFWYTMCPA 185  
QY 468 CEKLYGGAVP 477  
DB 186 CEKLYGGAVP 195

Search completed: June 17, 2003, 16:43:24  
Job time : 94.2277 secs